

# **Predicción de fusiones y adquisiciones a través de técnicas de aprendizaje automático**

**Predicting M&A with machine learning techniques**

Trabajo Fin de Grado

Doble grado en Ingeniería Informática y Administración de Empresas

Autor: D. Sebastián Freire García.

Tutor Administración de Empresas: D. Iván Blanco Sánchez

Tutor Ingeniería Informática: D. José Luis Torrecilla Noguerales

Universidad Carlos III de Madrid

Colmenarejo, 22 de junio de 2.017



## **Agradecimientos**

A mis tutores, José Luis Torrecilla Nogueras e Iván Blanco Sánchez por la ayuda y el tiempo prestados.

A mi familia, por confiar en mí y darme apoyo en todo momento.

A mis amigos, a los de aquí y a los de allí, porque sin ellos no sería quien soy.

A todos, gracias.



*“Si usted desea triunfar, tiene que duplicar su tasa de fracasos” Tom Watson jr.*



## Resumen

Este trabajo Fin de Grado estudia la predicción de fusiones y adquisiciones de empresas a través de técnicas de aprendizaje automático. La predicción de estas operaciones mediante estas técnicas resulta un problema interesante ya que todavía no se ha abordado correctamente y por lo tanto está sujeta a mejoras. Para la predicción disponemos de un conjunto de observaciones con datos financieros de empresas americanas comprendido entre 1994 y 2014. El estudio se lleva a cabo con diferentes clasificadores con el fin de evaluar los resultados desde diferentes enfoques. Los análisis muestran resultados bastante pesimistas en todos los algoritmos de clasificación, no obstante, se estudian las estrategias de inversión aplicables y como obtener rendimientos positivos de una fusión o adquisición.

**Palabras clave:** Adquisición, Aprendizaje Automático, Arbitraje, Clasificación, Fusión, R, Observación, Panel, Predicción.





## Tabla de Contenidos

<b>CAPÍTULO 1 INTRODUCCIÓN.....</b>	<b>1</b>
1.1    MOTIVACIÓN, OBJETIVOS Y LIMITACIONES .....	1
1.2    ESTRUCTURA DEL DOCUMENTO .....	3
<b>CAPÍTULO 2 REVISIÓN DE LA LITERATURA.....</b>	<b>5</b>
2.1    DEFINICIÓN Y MOTIVOS PARA LA FUSIÓN O ADQUISICIÓN .....	5
2.1.1 <i>Motivos considerables para la absorción.....</i>	<i>6</i>
2.1.2 <i>Motivos dudosos para la absorción .....</i>	<i>8</i>
2.2    CICLOS DE FUSIONES Y ADQUISICIONES.....	9
2.3    PRIMAS Y RENDIMIENTOS EN LOS ANUNCIOS DE ABSORCIÓN.....	11
2.4    PREDICCIÓN DE FUSIONES Y ADQUISICIONES.....	14
2.4.1 <i>Factores de predicción e hipótesis usados en estudios empíricos.....</i>	<i>15</i>
2.5    APRENDIZAJE AUTOMÁTICO.....	16
2.5.1 <i>Áreas de aplicación del aprendizaje automático .....</i>	<i>19</i>
2.5.2 <i>Aprendizaje automático en finanzas.....</i>	<i>19</i>
<b>CAPÍTULO 3 METODOLOGÍA. ....</b>	<b>21</b>
3.1    COMPRENSIÓN DEL NEGOCIO.....	21
3.2    COMPRESIÓN DE LOS DATOS .....	22
3.3    PREPARACIÓN DE LOS DATOS .....	23
3.4    MODELADO .....	23
3.5    EVALUACIÓN .....	23
3.6    DESPLIEGUE .....	24
<b>CAPÍTULO 4 DATOS. ....</b>	<b>25</b>
4.1    COMPRENSIÓN DE LOS DATOS.....	25
4.1.1 <i>Estructura de los datos .....</i>	<i>25</i>
4.1.2 <i>Calidad de los datos.....</i>	<i>27</i>
4.1.3 <i>Variable dependiente.....</i>	<i>29</i>
4.1.4 <i>Estadísticas descriptivas .....</i>	<i>29</i>
4.2    PREPARACIÓN DE LOS DATOS .....	31

4.2.1	<i>Selección de variables</i>	31
4.2.2	<i>Formateo de los datos</i>	35
4.2.3	<i>Winsorización y normalización</i>	35
<b>CAPÍTULO 5 ESTUDIO EMPÍRICO.</b>		<b>37</b>
5.1	PROBLEMA DE CLASIFICACIÓN	37
5.2	TÉCNICAS DE CLASIFICACIÓN	39
5.2.1	<i>Análisis discriminante lineal</i>	39
5.2.2	<i>k - vecinos cercanos</i>	41
5.2.3	<i>Máquinas de vectores soporte</i>	42
5.2.4	<i>Random Forest</i>	45
5.3	RECURSOS UTILIZADOS	46
5.3.1	<i>Equipo</i>	46
5.3.2	<i>R como lenguaje de programación</i>	46
5.3.3	<i>Microsoft Excel</i>	48
5.4	MODELO DE CLASIFICACIÓN	49
5.4.1	<i>Conjunto de entrenamiento y de test</i>	49
5.4.1	<i>Segmentación</i>	51
5.5	MODELO DE EVALUACIÓN	53
<b>CAPÍTULO 6 EVALUACIÓN DE RESULTADOS.</b>		<b>57</b>
6.1	RESULTADOS	57
6.2	EVALUACIÓN DE RESULTADOS	59
<b>CAPÍTULO 7 EXPLOTACIÓN DEL MODELO.</b>		<b>63</b>
7.1	ARBITRAJE	63
7.1.1	<i>Estrategias de arbitraje en operaciones de fusión y adquisición</i>	65
7.2	ESTRATEGIAS DE INVERSIÓN	67
7.2.1	<i>Arbitraje en largo</i>	68
7.2.2	<i>Arbitraje con empresas correlacionadas</i>	69
7.2.3	<i>Evaluación estrategias</i>	70
<b>CAPÍTULO 8 PLANIFICACIÓN Y PRESUPUESTO</b>		<b>71</b>
8.1	PLANIFICACIÓN	71

8.1.1	<i>Asignación de horas estimadas.....</i>	<i>71</i>
8.2	PRESUPUESTO.....	74
<b>CAPÍTULO 9 MARCO REGULADOR .....</b>		<b>77</b>
9.1	OPERATIVA .....	77
9.1.1	<i>Legislación .....</i>	<i>77</i>
9.1.2	<i>Formas en la adquisición .....</i>	<i>78</i>
9.1.3	<i>Contabilidad.....</i>	<i>78</i>
9.1.4	<i>Impuestos.....</i>	<i>79</i>
9.2	DESPLIEGUE .....	79
<b>CAPÍTULO 10 CONCLUSIONES Y TRABAJO FUTURO .....</b>		<b>81</b>
10.1	CONCLUSIONES .....	81
10.2	TRABAJO FUTURO .....	82
<b>LISTA DE REFERENCIAS.....</b>		<b>85</b>
<b>APÉNDICE .....</b>		<b>91</b>
APÉNDICE 1: DETALLE DEL PANEL DE DATOS .....		91
APÉNDICE 2: FLUJO DEL PROYECTO.....		96
APÉNDICE 3: APLICACIÓN DE LAS TÉCNICAS DE CLASIFICACIÓN .....		97
APÉNDICE 4: COTIZACIONES SIMULADAS .....		102
<i>Estrategia en largo.....</i>		<i>102</i>
<i>Estrategia con empresas correlacionadas.....</i>		<i>105</i>
<b>ANEXO .....</b>		<b>108</b>
ANEXO 1: PALEPU (1986) .....		108
<b>INTRODUCTION CHAPTER.....</b>		<b>111</b>
MOTIVATION, GOALS AND LIMITATIONS .....		111
DOCUMENT STRUCTURE .....		113
<b>LITERATURE REVIEW CHAPTER .....</b>		<b>115</b>
DEFINITIONS AND CAUSES FOR M&A.....		115
MACHINE LEARNING.....		116

<b>METHODOLOGY CHAPTER .....</b>	<b>119</b>
BUSINESS UNDERSTANDING .....	119
DATA UNDERSTANDING.....	120
DATA PREPARATION .....	120
MODELING .....	120
EVALUATION .....	121
DEPLOYMENT.....	121
<b>CONCLUSIONS AND FUTURE WORK CHAPTER.....</b>	<b>123</b>
CONCLUSIONS .....	123
FUTURE WORK .....	124

## Lista de tablas

TABLA 1 – RENDIMIENTOS DE LAS ACCIONES DE LA EMPRESA OBJETIVO EN RELACIÓN AL ANUNCIO DE UNA OPERACIÓN DE ABSORCIÓN..	13
TABLA 2 - ESTRUCTURA DEL PANEL DE DATOS DETALLE EMPRESA.....	26
TABLA 3 – NÚMERO DE OPERACIONES ‘0’ Y ‘1’ CONTEMPLADAS EN EL ESTUDIO. 27	
TABLA 4- INCIDENCIAS DETECTADAS EN EL PANEL INICIAL DE DATOS. SE DIVIDEN EN DOS GRUPOS GC1 Y GC2..	28
TABLA 5- EJEMPLO SOBRE LA PRESENCIA DE VALORES ATÍPICOS EN VARIABLES RELEVANTES DEL ESTUDIO.....	28
TABLA 6- ESTADÍSTICOS DESCRIPTIVOS.....	30
TABLA 7- NÚMERO DE OBSERVACIONES EN FUNCIÓN DEL FACTOR MERGER. ....	50
TABLA 8 – CONJUNTOS DE ENTRENAMIENTO Y DE TEST..	51
TABLA 9 - MATRIZ DE CONFUSIÓN..	53
TABLA 10 - MATRIZ DE CONFUSIÓN PARA LOS DATOS DE TEST APLICADOS AL ALGORITMO DE CLASIFICACIÓN LDA.....	57
TABLA 11 - MATRIZ DE CONFUSIÓN PARA LOS DATOS DE TEST APLICADOS AL ALGORITMO DE CLASIFICACIÓN KNN .....	58
TABLA 12 - MATRIZ DE CONFUSIÓN PARA LOS DATOS DE TEST APLICADOS AL ALGORITMO DE CLASIFICACIÓN SVM LINEAR.....	58
TABLA 13 - MATRIZ DE CONFUSIÓN PARA LOS DATOS DE TEST APLICADOS AL ALGORITMO DE CLASIFICACIÓN SVM GAUSSIANO. ....	58
TABLA 14 - MATRIZ DE CONFUSIÓN PARA LOS DATOS DE TEST APLICADOS AL ALGORITMO DE CLASIFICACIÓN RANDOM FOREST. ....	58
TABLA 15 – RESUMEN MÉTODOS DE CLASIFICACIÓN UTILIZADOS. ....	59
TABLA 16 – ESTRATEGIA EN LARGO PARA TRES EMPRESAS DE DIFERENTES SECTORES MARCADAS COMO ‘1’. L.....	68
TABLA 17- ESTRATEGIA CON EMPRESAS CORRELACIONADAS PARA 3 EMPRESAS DE DIFERENTES SECTORES..	69
TABLA 18 – DETALLES HORAS ESTIMADAS.....	73
TABLA 19 – DIAGRAMA GANTT PLANIFICACIÓN.....	74
TABLA 20 – PRESUPUESTO COSTE DEL PERSONAL. ....	75
TABLA 21 – PRESUPUESTO COSTE DEL EQUIPAMIENTO.....	75
TABLA 22 – PRESUPUESTO COSTE DEL SOFTWARE. ....	75
TABLA 23 – PRESUPUESTO DEL COSTE TOTAL DEL PROYECTO. ....	76
TABLA 24 – CARÁCTERÍSTICAS DEL PANEL DE DATOS.....	92
TABLA 25 – RESULTADOS PARA (PALEPU, 1986).....	108



## Lista de figuras

ILUSTRACIÓN 1 – HISTÓRICO DE OPERACIONES DE ABSORCIÓN EN EL PERIODO 1980 – 2015.....	10
ILUSTRACIÓN 2 – AGENTE Y SU ENTORNO.....	18
ILUSTRACIÓN 3- ETAPAS DE LA METODOLOGÍA CRISP – DM .....	22
ILUSTRACIÓN 4 – EJEMPLO DE UN DISCRIMINANTE LINEAR PARA DOS CLASES .....	40
ILUSTRACIÓN 5 – K- VECINOS CERCANOS. ....	42
ILUSTRACIÓN 6 – HIPERPLANO SVM .....	44
ILUSTRACIÓN 7 – EJEMPLO DE PROBLEMA DE CLASIFICACIÓN QUE NO ES LINEALMENTE SEPARABLE .....	45
ILUSTRACIÓN 8 - SEGMENTACIÓN.....	52
ILUSTRACIÓN 9 – REPRESENTACIÓN DE SENSIBILIDAD (RECALL) Y PRECISIÓN. ....	55
ILUSTRACIÓN 10 – FLUJO DE PROGRAMACIÓN DEL PROYECTO.....	96
ILUSTRACIÓN 11 – ARBITRAJE EN LARGO EMPRESA A. ....	102
ILUSTRACIÓN 12 - ARBITRAJE EN LARGO EMPRESA B.....	103
ILUSTRACIÓN 13 - ARBITRAJE EN LARGO EMPRESA C.....	104
ILUSTRACIÓN 14 – ESTRATEGIA CON EMPRESAS CORRELACIONADAS EMPRESA A .....	105
ILUSTRACIÓN 15 - ESTRATEGIA CON EMPRESAS CORRELACIONADAS EMPRESA B.. ....	106
ILUSTRACIÓN 16 - ESTRATEGIA CON EMPRESAS CORRELACIONADAS EMPRESA C, .....	107





## **Capítulo 1**

### **Introducción**

#### **1.1 Motivación, objetivos y limitaciones**

Las fusiones y adquisiciones de empresas representan, cada vez más, un factor importante para los negocios desarrollados en un entorno de alta competitividad y es una gran herramienta que las compañías utilizan para mantener o incluso extender su poder en el mercado y sus ventajas competitivas.

El avance de las técnicas de computación y transmisión de información ha generado la creación de inmensas cantidades de información. Todavía, mucha de esta información aún no ha sido procesada y continúa almacenada sin ser explotada, siendo únicamente datos. Como existe una gran cantidad de datos almacenados, la información potencial que se puede obtener a partir de ellos es presumiblemente alta. A través de técnicas de aprendizaje automático se pueden analizar y modelar los datos, detectando patrones y de este modo recabar información útil.

Demostrado por diferentes estudios Jarrell & Poulsen (1989) y Andrade, Mitchell, & Stafford (2001) es bien sabido que los accionistas de la empresa adquirida reciben altos beneficios tras el anuncio de compra de la empresa. Por lo tanto, encontrar formas de identificar estas compañías antes de que una oferta de compra se anuncie es un problema de gran interés para la industria financiera. Sin embargo, este es un problema complejo que aún no tiene solución.

## 1.INTRODUCCIÓN

Varios estudios han propuesto modelos para la predicción de operaciones de fusión y adquisición basados en información pública orientados principalmente empresas del mercado estadounidense Palepu (1986), Meador, Pamela, & Gayle (1996), Barnes (2000), Ragothaman, S., Naik, B., & Ramakrishnan (2003). La mayoría de estos modelos utilizan como técnica de predicción el análisis de factores, análisis de discriminante lineal o modelos de regresión logarítmicos.

El objetivo de este TFG es identificar factores de influencia y comprobar su viabilidad para operaciones de fusión y adquisición de empresas en el mercado americano. Para ello se implementarán técnicas de aprendizaje automático que a partir de datos financieros tratarán de predecir si una empresa será adquirida al final de un trimestre o no. A partir de los resultados se analizará el potencial para generar rendimientos positivos utilizando estrategias de inversión propias de las operaciones de absorción.

Este TFG intenta contribuir a los estudios existentes en tres puntos. Primero, el uso de técnicas de aprendizaje automático para la predicción de fusiones y adquisiciones. Segundo, la viabilidad del uso de datos totalmente financieros y públicos. Tercero, la aplicabilidad de los resultados obtenidos a un escenario real.

No obstante, se presentan diferentes limitaciones para la realización de un modelo de predicción único, válido para todas las empresas. Las técnicas de aprendizaje automático únicamente procesarán datos financieros que, aunque aportan un valor importante a la predicción, no abarcan todos los motivos por los que se producen las absorciones. Existen otras variables no contempladas en el estudio como son las variables de mercado (Meador et al., 1996), las variables macroeconómicas (Rossi & Volpin, 2004), tecnológicas (Yang, 2014) y variables relacionadas con la estructura de control (Ambrose & Megginson, 2011) que también presentan influencias en la decisión de una operación de absorción. En

relación con los datos usados en el estudio, aunque han sido obtenidos de fuentes fiables (*Compustat* y *SDC Platinum*) no se ha podido verificar la validez de todos ellos y se han asumido como correctos para la predicción.

## **1.2 Estructura del documento**

Una vez expuestos los objetivos principales que pretende el desarrollo de este trabajo, la estructura del documento se puede dividir en tres bloques; a los cuales los precede la introducción y los sigue el apartado de conclusiones.

El primer bloque (Capítulo 2) contiene los fundamentos teóricos en materia de fusiones y adquisiciones de empresa. Presenta los motivos que llevan a dos empresas a fusionarse, las influencias de las tendencias macroeconómicas que existen en este tipo de operaciones y los modelos de predicción utilizados. También presenta de una manera general lo que es el aprendizaje automático y sus áreas de aplicación.

El segundo bloque (Capítulo 3,4,5,6) describe la elaboración del modelo de predicción. Metodología empleada, datos utilizados, elaboración del modelo y evaluación del mismo. En este bloque se describen también fundamentos teóricos para poder entender las decisiones tomadas.

El tercer bloque (Capítulo 7,8,9) comprende la explotación del modelo de predicción, presupuesto y planificación y el marco regulador que le afecta directamente.

## 1.INTRODUCCIÓN

## Capítulo 2

### Revisión de la literatura

En esta sección realiza una revisión de la literatura relacionada con las fusiones y adquisiciones de empresas (M&A) y los distintos métodos para predecirlas. Se abordarán temas relacionados con este tipo de operaciones como valor añadido, motivos y tipos. En cuanto a los métodos de predicción se hará un resumen de los procedimientos empleados y por último se explicará el enfoque del aprendizaje automático orientado a sus áreas de aplicación.

#### 2.1 Definición y motivos para la fusión o adquisición

El aumento del ritmo de las operaciones de fusión y adquisición en Estados Unidos en los últimos años ha sido notable. Brealey, Myers, & Allen (2013) afirma:

2006 ha sido un año récord en operaciones de fusión y adquisición, las compañías han estado involucradas en acuerdos por un total de \$1.5 trillones en Estados Unidos y más de \$0.5 trillones en Europa. Las compañías asiáticas, que habitualmente han estado rezagadas en este tipo de operaciones, registraron \$59 billones en acuerdos. (p.824)

El número de este tipo de operaciones ha aumentado y en 2015 ha sido de \$5.0 trillones. Durante estos periodos de intensa actividad, la dirección de las empresas emplea cantidades significativas de tiempo intentando prever si una empresa adquirirá a otra.

## 2. REVISIÓN DE LA LITERATURA

Comúnmente nos referimos a los términos de fusión o adquisición prácticamente de la misma manera, pero estos son diferentes. Una fusión surge cuando dos compañías deciden unirse para formar una sola empresa. Por otro lado, una adquisición, por regla general, existe cuando una empresa de gran envergadura decide comprar una empresa más pequeña. Al tratarse de operaciones diferentes siguen etapas y restricciones diferentes, pero el resultado final es prácticamente el mismo a efectos del comportamiento de las compañías en bolsa. Por lo tanto, a lo largo de este estudio no se realizará diferencias entre estas operaciones agrupándolas bajo el término de operaciones de *absorción* y analizándolas bajo los supuestos enunciados por Tobergte & Curtis (2013) que usa el término absorción para cualquier adquisición o control corporativo a través de la compra de acciones de una empresa objetivo, independientemente de que la operación se haya realizado a través de un acuerdo o una oferta pública.

Las operaciones de absorción principalmente centran todo su atractivo al resultado económico de combinar dos empresas, pero también están relacionadas con cuál de las dos consigue la gestión de la compañía resultante. En cualquier operación de fusión se considera que existe una firma protagonista considerada *adquisidor* o *acquirer* y otra considerada como *objetivo* o *target*, en la que normalmente la directiva de la firma adquirida deja sus puestos después de la operación.

El mercado de fusiones y adquisiciones actualmente se considera como parte de un mercado mayor *market for corporate control* que incluye privatizaciones, nacionalizaciones y compras de empresas en concurso de acreedores entre otros. Estas operaciones, que surgen por otros motivos y tienen otras consecuencias, no serán analizadas durante el estudio.

### 2.1.1 Motivos considerables para la absorción

Una operación de absorción únicamente añade valor si las dos compañías pasan a valer más juntas que de forma separada. Por lo tanto, las empresas cuando realizan este tipo de operaciones se basan en buscar sinergias y de esta forma encontrar beneficios en trabajar de forma conjunta. De acuerdo con Brealey et al. (2013), las sinergias pueden ser:

- Economías de escala: muchas operaciones de absorción son planeadas para reducir costes y de esta forma conseguir economías de escala. La fusión entre las empresas Duke Energy and Progress Energy (Duke Energy Corporation, 2012) se realizó previendo una reducción en los costes de \$1.6 billones durante cinco años.
- Economías de integración vertical: con estas operaciones se busca ganar control sobre el proceso productivo integrándose hacia delante o hacia atrás. Un método para conseguir esto es fusionarse o adquirir la empresa proveedora o cliente. De todos modos, este tipo de operaciones han dejado de ser comunes ya que las empresas han encontrado más eficiente la externalización de muchos servicios y métodos de producción.
- Recursos complementarios: muchas pequeñas firmas han sido adquiridas por firmas más grandes ya que estas tienen la capacidad de proveer pequeños recursos complementarios y muy específicos para el desarrollo de la actividad de las empresas grandes. Un ejemplo de este tipo de sinergia puede ser la compra de Instagram y Whatsapp por parte de Facebook, estas dos empresas más pequeñas le proporcionaban a la tercera un servicio a usuarios de redes sociales que Facebook no abarcaba.

## 2. REVISIÓN DE LA LITERATURA

- Excedente de fondos: estas sinergias se producen en mercados maduros en los que la empresa genera una alta cantidad de beneficios, pero tiene pocas posibilidades de realizar una inversión beneficiosa. En vez de distribuir el beneficio entre los accionistas, muchos directores buscan comprar acciones de otra compañía. De este modo, compañías con un excedente de liquidez y escasez de buenas oportunidades de inversión a menudo convierten las adquisiciones financiadas con liquidez como una buena forma de reubicar su capital.
- Eliminar ineficiencias: existen empresas con oportunidades sin explotar en cuanto a la optimización de costes e incremento de las ventas y beneficios. Estas empresas son candidatos naturales para la adquisición por parte de otras firmas con una mejor gestión.
- Consolidación de la industria: la mayor oportunidad para mejorar la eficiencia viene de industrias en las que operan muchas firmas con mucha capacidad. Estas condiciones suelen propiciar olas de fusiones y adquisiciones, ya que fuerza a las empresas a recortar capacidad y recursos, y liberar capital para reinvertirlo posteriormente de nuevo en la empresa. Un ejemplo de este tipo de sinergias serían las operaciones de fusión bancaria que se dieron tras la crisis de 2008, donde muchas entidades se vieron obligadas a fusionarse para poder seguir ejerciendo su actividad económica.

### 2.1.2 *Motivos dudosos para la absorción*

A parte de las sinergias descritas con anterioridad, las cuales poseen un sentido económico, existen otros motivos de carácter dudoso por los que las empresas realizan operaciones de fusión o adquisición (Brealey et al., 2013):



- Diversificación: los directores de empresas con alta liquidez prefieren destinar esa liquidez a adquisiciones antes que repartirlo en dividendos. El problema viene cuando esta diversificación se produce desde mercados en desarrollo a mercados en madurez.
- Incrementar los ingresos por acción *The Bootstrap Game*: se produce en adquisiciones que no ofrecen evidentes beneficios económicos pero que sin embargo producen varios años de incremento en los ingresos por acción.
- Reducción en los costes de financiación: estas se producen bajo la asunción de que cuando dos firmas se unen, la compañía resultante podrá financiarse a menores tasas de interés de lo que haría de forma separada. Cuando dos empresas están separadas, una no responde por la deuda de la otra, en el caso de que estén fusionadas tienen más medios para responder a la deuda y por lo tanto disminuye su tasa de interés.

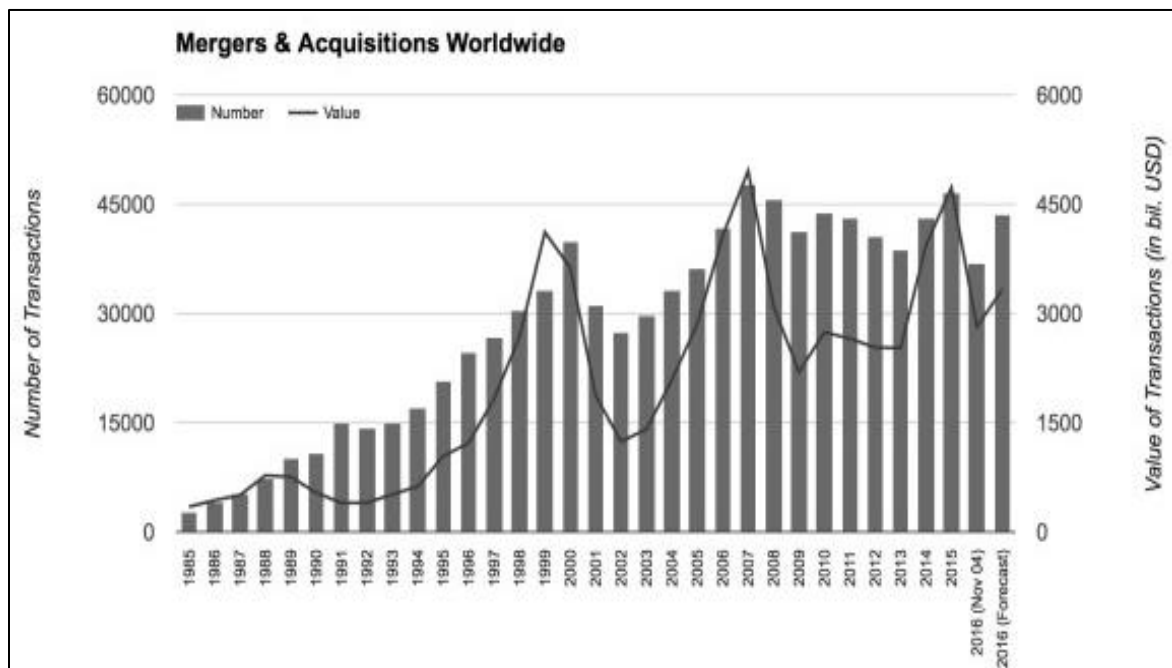
## 2.2 Ciclos de fusiones y adquisiciones

Como señalan Mitchell & Mulherin (1996) la actividad de fusiones y adquisiciones está guiada por motivos macroeconómicos y cambios en la industria que activan comienzos de alta actividad. En su estudio destacan seis incrementos de fusiones y adquisiciones a lo largo de la historia [*Ilustración 1*].

El primer incremento fue entre 1879-1904 fue caracterizado por fusiones horizontales con la posterior generación de monopolios. El segundo incremento (1916-1929) se describe como un aumento de actividad que incluye varias adquisiciones verticales las cuales dieron lugar a una gran cantidad de oligopolios debido a la imposibilidad de formar monopolios,

## 2. REVISIÓN DE LA LITERATURA

regulado por la ley antimonopolio del *Clayton Act de 1914*. Tras el *crash* de la bolsa del 29 el número de operaciones de absorción resultó reducido. El tercero fue en el periodo de 1965 a 1969 caracterizado por fusiones de conglomerados con carteras de productos variadas. El cuarto incremento (1984-1989) destaca por involucrar altas cantidades de dinero en las operaciones, se llegaron a invertir un total de \$100 millones. Después de la recesión de 1990 y hasta casi los 2000, surge el quinto ciclo en la historia de las fusiones y adquisiciones donde las empresas se preocuparon en la creación de acuerdos estratégicos. La última ola (2003-2007) es característica por las operaciones internacionales, muchas de integración horizontal y operaciones de compra apalancada. Esto puede ser debido a un periodo de ratios de interés bajo y el aumento de transacciones a nivel mundial.



*Ilustración 1 - Histórico de operaciones de absorción en el periodo 1980 – 2015. Las barras representan el número de operaciones en cada año y la línea indica el valor de las operaciones. Fuente: Thompson Analytics.*

### **2.3 Primas y rendimientos en los anuncios de absorción**

En las operaciones de absorción, la empresa adquisidora normalmente paga una prima a los accionistas de la empresa objetivo. Asumiendo mercados de capital eficientes, los precios de las acciones de la empresa adquisidora y de la empresa adquirida deberían ajustarse instantáneamente. De esta forma sería posible unir toda la información pública sobre la transacción (Fama & French, 1992), esto incluye la prima, la forma de pago y cualquier tipo de información relevante para los inversores. De todos modos, esto no ocurre instantáneamente ya que los mercados no muestran una eficiencia perfecta, así que a partir de estas operaciones se generan diferencias en los precios que los inversores tratan de explotar.

Estudios empíricos han mostrado que los accionistas de la empresa objetivo reciben rendimientos sustanciales en un periodo de tiempo dado alrededor del día del anuncio de una oferta de adquisición. Andrade et al. (2001) observan que los accionistas de la empresa adquirida reciben un incremento del 16% en el precio de sus acciones desde un día antes hasta un día después del anuncio de la operación. Cuando el periodo se extiende a 20 días antes del anuncio hasta la fecha en la que se cierra la operación el estudio indica que el incremento en el precio de las acciones pasa a ser de un 23.8%.

Varios estudios Langetieg (1978), Bradley & Kim (1988), Jarrell & Poulsen (1989), Schwert et al. (1996), Andrade et al. (2001), Horning et al. (2014) confirman una subida del precio de las acciones de la empresa objetivo hasta el momento del anuncio de adquisición [Tabla 1]. Según las investigaciones llevadas a cabo por Langetieg (1978) las empresas anunciadas como objetivo aumentan el valor de sus acciones un 10,63% desde cuatro meses antes hasta el momento de su adquisición. Bradley & Kim (1988) argumentan que las empresas anunciadas como objetivo de una OPA desde cinco días antes hasta cinco

## 2. REVISIÓN DE LA LITERATURA

días después aumentan su valor en un 37,71%. Jarrell & Poulsen (1989) se fijan en empresas involucradas con empresas americanas donde desde veinte días antes hasta diez días después de la operación la empresa marcada como objetivo aumenta un 28,99% su cotización. Schwert et al. (1996) confirma que para operaciones que terminan con una OPA satisfactoria el porcentaje de subida de las acciones de la empresa objetivo se traduce en un 30,1%.

Diferentes son los factores influyentes en el nivel de rendimiento de los accionistas de las empresas adquiridas- Estos se pueden clasificar en dos grupos: en primer lugar se encuentran los factores relacionados con la operación, que tienen relación con la participación de la empresa adquisidora, tipo de pago o si la operación es internacional o doméstica (Rossi & Volpin, 2004). En segundo lugar, están los factores no relacionados con la operación sino con la empresa objetivo, que no están influenciados por el tipo de adquisidor o las decisiones de este. Destacan entre estos algunos indicadores como el tamaño de la empresa (Melicher & Nielsen, 1977), los ratios de libros a mercado (Eckbo, 2009) y los de mercado a libros (Walkling & Edminster, 1985).

PREDICCIÓN DE FUSIONES Y ADQUISICIONES A TRAVÉS DE TÉCNICAS DE  
APRENDIZAJE AUTOMÁTICO



Estudio	Rendimiento	Periodo	Marco duración estudio (días)	Detalles del estudio
(Langetieg, 1978)	+10,63%	1929-1969	(-126,0)	- Muestra 149 empresas - Transacciones USA - 90% muestra después 1950
(Bradley & Kim, 1988)	+31,77%	1963-1984	(-5,+5)	- Muestra 236 empresas - Ambas empresas estadounidenses y cotizando - Sólo OPAs exitosas
(Jarrell & Poulsen, 1989)	+28,99%	1963-1986	(-20,+10)	- Muestra 236 empresas - Una de las dos empresas cotiza y es estadounidense
(Schwert et al., 1996)	+30,1%	1975-1991	(-42,+126)	- Muestra 1.174 empresas - Ambas empresas estadounidenses y cotizando - Sólo OPAs exitosas
(Andrade et al., 2001)	+16,0% +23,8%	1973-1998 1973-1998	(-1,+1) (-20, Cierre)	- Muestra 3.688 empresas - Ambas empresas estadounidenses y cotizando

*Tabla 1 – Rendimientos de las acciones de la empresa objetivo en relación al anuncio de una operación de absorción. La tabla proporciona un resumen de los estudios empíricos más importantes relacionados con el anuncio de operaciones de absorción en Estados Unidos.*

## 2. REVISIÓN DE LA LITERATURA

A diferencia de en las empresas objetivo, no se puede asegurar un retorno para los accionistas de la empresa adquirente. Literatura enfocada al estudio de las operaciones de absorción en el mercado estadounidense muestra pequeñas subidas y bajadas en el precio de la acción del adquirente (Andrade et al., 2001; Jarrell & Poulsen, 1989; Langetieg, 1978; Schwert et al., 1996) sin la misma significación estadística que para las fluctuaciones en la empresa objetivo. R.F. Bruner (2004) en su libro *Applied merger and acquisitions* muestra que los rendimientos (negativos, positivos o nulos) están distribuidos de una manera similar a lo largo del histórico de fusiones y adquisiciones en el mercado estadounidense y siendo la mayoría de retornos entre un -5% y un +5%.

### **2.4 Predicción de fusiones y adquisiciones**

La predicción de fusiones y adquisiciones usando datos de dominio público ha sido objeto de varios estudios desde el inicio de 1970. Muchos de estos primeros estudios estaban motivados por el beneficio potencial que se podía obtener de las predicciones de futuras adquisiciones. La mayoría de las técnicas de predicción empleadas se centran en el estudio de la empresa objetivo, dejando casi siempre de lado las empresas que realizan la oferta. El estudio de referencia sobre las buenas prácticas en la realización de estos modelos de predicción está diseñado por Palepu (1986) que reseñaba que el acierto de la predicción sobre el mercado no era fiable muchas veces hasta tres meses antes de la fecha de anuncio de la adquisición.

En las tres últimas décadas, se han realizado estudios empíricos que a través de modelos estadísticos (usando técnicas de regresión logística) también han intentado predecir operaciones de fusión y adquisición. La mayoría de estudios realizados están orientados al mercado de Estados Unidos, pero también en otros como Canadá, Europa (Ambrose & Megginson, 2011; Cremers, Nair, & John, 2009; Dietrich & Sorenson, 1984; Palepu, 1986).

También se utilizado técnicas de discriminante lineal (Barnes, 2000; Ragothaman, S., Naik, B., & Ramakrishnan, 2003; Simkowitz & Monroe, 1971; Wansley, 1984), análisis de factores (Stevens, 1973), modelos condicionales de regresión logarítmica (Tsagkanos, A., Georgopoulos, A., & Siriopoulos, 2007) y análisis probit (Harris, Stewart, Guilkey, & Willard T., 1982).

#### 2.4.1 Factores de predicción e hipótesis usados en estudios empíricos

Como se ha comentado con anterioridad, existen diferentes factores que han sido objeto de estudio como indicadores de operaciones de absorción. A continuación procederemos a enumerarlos y describirlos.

- Factores microeconómicos: Palepu (1986) consolidó una base para la predicción de fusiones y adquisiciones estableciendo una serie de hipótesis que posteriormente serían utilizadas por la mayoría de investigadores:
  1. *Inefficient management hypotesis*: las empresas con gestión ineficiente es probable que sean adquiridas.
  2. *Growth-resource mismatch hypotesis*: las empresas con una discordancia entre su crecimiento y recursos financieros es probable que sean adquiridas.
  3. *Industry disturbance hypotesis*: las empresas que operen en un sector sujeto a ‘turbulencias económicas’ es probable que sean adquiridas.
  4. *Size hypotesis*: la probabilidad de adquisición se reduce con el tamaño de la empresa.
  5. *Market-to-book hypotesis*: las empresas cuyo valor en el mercado sea menor que su valor en libros es probable que sean adquiridas.
  6. *Price-earnings hypotesis*: las empresas con ratios bajos de P/E es probable que sean adquiridas.

## 2. REVISIÓN DE LA LITERATURA

Existen otros estudios que añaden a las hipótesis descritas con anterioridad un factor financiero determinante integrado en varios estudios empíricos como es el *leverage* o *apalancamiento* (Cremers et al., 2009).

Ambrose & Megginson (2011) analizaron la influencia de variables relacionadas con la estructura de control y con las estructuras de defensa (*poison pill*). Por ejemplo, si el accionariado de las empresas lo forman inversores institucionales o no donde se puede observar que el comportamiento de las empresas cambia cuando son propiedad de inversores institucionales.

- Factores macroeconómicos: un tipo de variables que los investigadores no han cubierto demasiado en los modelos de predicción. Robert F. Bruner (2011) argumenta que las actividades de fusión y adquisición están generalmente influenciadas por factores macroeconómicos y del sector (PIB, tasas de interés, tipos de cambio, política fiscal, política monetaria...), mercados de capital y de deudas (interés de la deuda, ratings, múltiplos, betas...) y comportamiento del consumidor (elasticidad-precio de la demanda).

### 2.5 Aprendizaje automático

El aprendizaje automático es un subcampo de las ciencias de la computación y una rama de la inteligencia artificial que tiene como objetivo la elaboración de técnicas que permitan a los ordenadores aprender sin ser programados explícitamente (Rouse, 2017). A partir de información proporcionada en forma de ejemplos, el aprendizaje automático es capaz de crear programas capaces de generalizar comportamientos.

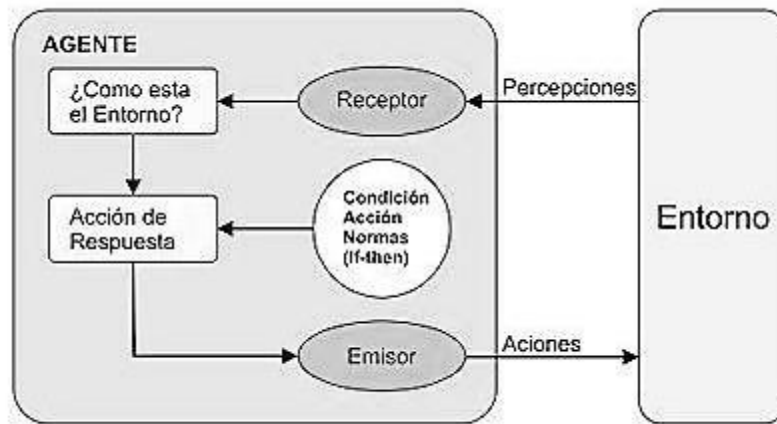


Las técnicas de aprendizaje automático intentan resolver problemas del dominio de la estadística que presentan una alta complejidad computacional, de este modo surgen diferentes disciplinas de esta materia como el análisis de datos o el estudio de la complejidad computacional.

Para entender de donde proviene del aprendizaje automático dentro del área de la inteligencia artificial es necesario primero entender los conceptos de agente, entorno, sensor y actuador. Russel & Norvig (2010): “Un agente es cualquier cosa que se puede ver o percibir en su entorno o a través de sensores y que interactúa con dicho entorno a través de actuadores”. Un ejemplo aplicado a la robótica sería el caso de un robot que tenga cámaras y detectores de infrarrojos como sensores y varios motores como actuadores y actúe en una determinada localización. Una vez establecidos estos cuatro conceptos el objetivo de la inteligencia artificial es conseguir que el agente se comporte de manera racional, es decir, que por cada posible secuencia de acciones, el agente seleccione una acción que maximice su rendimiento.

El comportamiento del agente se ve afectado por lo que perciba en el entorno por medio de sus sensores, a esto se le denomina problema de entorno. Este concepto considera los problemas que se presentan a los agentes racionales y que son independientes de ellos y a los que estos deben aportar las soluciones [*Ilustración 2*].

## 2. REVISIÓN DE LA LITERATURA



*Ilustración 2 – Agente y su entorno. Los agentes interactúan con el entorno mediante sensores y actuadores y su comportamiento viene determinado por el estado del entorno.*

Tras hablar de los conceptos de “problema de entorno” y “agente racional” observamos que se pueden realizar diferentes clasificaciones en la inteligencia artificial dependiendo de estos dos conceptos. En relación al entorno podríamos hacer clasificaciones dependiendo de sus propiedades (total o parcialmente observable, estático o dinámico, determinista o estocástico ...) y en relación a los agentes en como han sido programados (utilidad, objetivos ...). Pero realmente la clasificación que nos interesa viene dada el tipo de datos obtenidos del entorno.

Para este estudio se utilizarán técnicas de aprendizaje automático que utilizan algoritmos de aprendizaje supervisado con un enfoque orientado a la clasificación. El aprendizaje supervisado consiste en realizar una predicción correcta de un conjunto de datos de entrada, con un algoritmo que previamente ha sido entrenado por un conjunto de datos de entrenamiento con una salida válida. Las predicciones pueden ser un valor continuo (como en los problemas de regresión) o una etiqueta de clase (clasificación) la cual presenta valores discretos.

### *2.5.1 Áreas de aplicación del aprendizaje automático*

El aprendizaje automático está presente en nuestro día a día en muchas de las relaciones que mantenemos con la tecnología. Por ejemplo, un viajero que quiera reservar unos billetes de avión lo puede realizar a través de una conversación guiada por un autómata con **reconocimiento del lenguaje** y un sistema que aprenda a gestionar el diálogo. Se han creado programas que **aprendan a jugar** mejor que muchos profesionales, uno de los primeros fue desarrollado por IBM y consiguió ganar a Kasparov en una partida de ajedrez. También aparece aprendizaje automático en la **detección automática de spam**, **detección de patrones en imágenes**, etc.

El aprendizaje automático también se utiliza en determinados sectores como una herramienta muy potente para la predicción de eventos. En el ámbito de la medicina existen nuevas líneas de investigación que a través de aprendizaje automático son capaces de diagnosticar enfermedades en pacientes (Di Deco Sampedro & Díaz García, 2012). En otros sectores como el eléctrico el aprendizaje automático se utiliza para realizar predicciones sobre los precios de la electricidad en las subastas.

### *2.5.2 Aprendizaje automático en finanzas*

El aprendizaje automático también ha encontrado un hueco en el mundo de las finanzas. Se puede destacar el uso técnicas de aprendizaje automático para las siguientes aplicaciones:

- Gestión de cartera de acciones: muchas compañías utilizan algoritmos para conseguir calibrar una cartera de acciones de acuerdo con los objetivos y riesgos que quiera tomar el gestor.

## 2. REVISIÓN DE LA LITERATURA

- Trading algorítmico: sus orígenes se remontan a 1970 e incluye el uso de sistemas complejos de inteligencia artificial para tomar decisiones de inversión extremadamente veloces (cada operación dura fracciones de segundo). Los sistemas algorítmicos realizan miles o millones de operaciones automáticamente en un día. Actualmente, estas operaciones se llevan a cabo con sistemas de negociación de alta frecuencia HFT (*high frequency trading*).
- Detección del fraude: combina la capacidad de computación del ordenador con el uso de la cantidad masiva de datos disponibles en internet. Estos sistemas de detección de fraude se basan en aprender y recalibrar sistemas básicos para detectar potenciales fallos de seguridad.
- Evaluación de préstamos y seguros: esta es la función a la que mejor se ha adaptado el aprendizaje automático en el mundo de las finanzas. Los algoritmos de aprendizaje automático pueden ser entrenados con grandes cantidades de datos de los usuarios (año, trabajo, estado civil) y de su estado financiero o de sus pólizas de seguros (si ha entrado en default, devolución del préstamo en plazo, si ha tenido un accidente de coche...). Se pueden distinguir comportamientos parecidos a través de algoritmos y continuamente realizar análisis para detectar otros comportamientos que puedan influenciar contratos futuros.

## Capítulo 3

### Metodología.

La metodología empleada en este proyecto está basada en CRISP-DM (Cross Industry Standard Process for Data Mining) (Marbán, Mariscal, & Segovia, 2009) ya que posee un enfoque global para resolver problemas de aprendizaje automático [Ilustración 3]. A partir de esta metodología crearemos un marco donde se establecen las diferentes partes de las que constará el experimento. Debido a que esta es una metodología usada en procesos de *minería de datos* y de *descubrimiento de conocimiento* y que el problema a resolver es un problema de *clasificación*, algunas partes se adaptarán al uso de técnicas de aprendizaje supervisado.

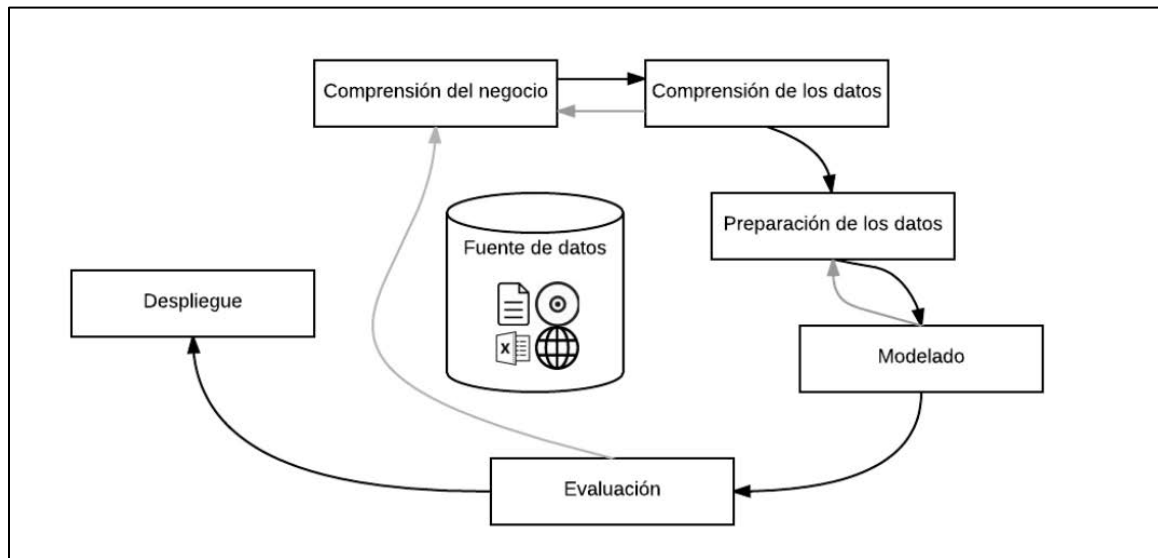
A continuación se describirán brevemente las etapas de esta metodología. Metodologías similares y más detalles de cada etapa pueden encontrarse en cualquier monografía sobre el tema (Marbán et al., 2009).

#### 3.1 Comprensión del negocio

Esta fase se centra en el entendimiento de los objetivos y requisitos del proyecto desde una perspectiva de negocio. En este caso, la creación de un portfolio de empresas que potencialmente puedan ser adquiridas al final de cuatrimestre.

### 3.METODOLOGÍA

Posteriormente se realiza una valoración del problema a resolver y se plantea el estudio desde un punto de vista orientado al aprendizaje automático. Se determinan los objetivos a alcanzar y finalmente se crea un plan para obtener los resultados deseados.



*Ilustración 3- Etapas de la metodología CRISP – DM*

#### **3.2 Comprensión de los datos**

La fase de comprensión de datos comienza con la recopilación de los datos de las fuentes de información. Continúa con un análisis preliminar, que consiste en obtener una primera imagen del tipo de datos del que se dispone, descripción y rasgos comunes. Posteriormente se realiza un análisis de la variable dependiente para observar sus características y descubrir cuáles son los factores que influyen en ella. También se realiza un análisis de calidad de los datos observando posibles datos atípicos, valores sin información y problemas relativos a la calidad para una futura subsanación. Finalmente, se realiza un simple análisis descriptivo de las variables más influyentes en el modelo con el objetivo de descubrir relaciones subyacentes en el conjunto de los datos.

### 3.3 Preparación de los datos

La preparación de los datos se centra en la adaptación de los datos iniciales para que posteriormente puedan ser utilizados de una manera adecuada. Esta adaptación se realiza en base al estudio de los datos realizado con anterioridad, obteniendo como resultado el conjunto de datos final. El proceso de preparación de los datos no tiene porque seguir un orden establecido ya que se realiza simultáneamente a la elaboración del modelo. Dentro de esta fase se incluye la selección, limpieza, construcción, integración y formateo de los datos.

### 3.4 Modelado

En esta fase se seleccionan las técnicas de modelado en función del tipo de problema a resolver. En este caso al tratarse de un problema de *clasificación* se utilizarán técnicas adecuadas. La elaboración del modelo de clasificación y evaluación son las partes fundamentales de esta fase.

Debido a que algunas técnicas poseen requisitos específicos en la estructura de los datos, esta fase no posee un orden preestablecido, siendo en algunos momentos necesario retroceder a pasos anterior para adecuar los datos a las técnicas empleadas en cada momento.

### 3.5 Evaluación

Una vez obtenidos los resultados del estudio empírico realizado con anterioridad, estos son evaluados desde una perspectiva de análisis de datos. Tras este análisis se realizará una revisión del proceso para posteriormente definir los pasos siguientes, es importante realizar una evaluación concienzuda de los resultados, modelo y de las decisiones tomadas para

### 3.METODOLOGÍA

asegurarse de que los resultados consiguen dar respuesta a los objetivos del problema de negocio planteado. Al final de esta fase se realiza una visión global de los puntos fuertes y débiles que presenta las técnicas aplicadas y posibles mejoras.

#### **3.6Despliegue**

En este estudio, la fase de despliegue indica la forma en la que se utilizaría el modelo de predicción y cómo se incluiría en los objetivos de negocio planteados.



## Capítulo 4

### Datos.

A lo largo de este capítulo se describen las decisiones tomadas y las operaciones realizadas con relación a los datos del problema. Consta de dos bloques principales orientados a la comprensión y al procesamiento.

#### 4.1 Comprensión de los datos

##### 4.1.1 Estructura de los datos

Para la realización del estudio se han empleado datos financieros y fiscales de empresas estadounidenses en un contexto histórico comprendido entre 1994 y 2014. La recopilación de datos se ha realizado a partir de dos fuentes diferentes, las bases de datos *Compustat* y *DC Platinum*. De *Compustat* se han obtenido detalles fiscales y financieros de la empresa, entre los que podemos destacar el capital social de la empresa (ceqq), ratio de rentabilidad financiera (roe) y apalancamiento (leverage). De *SDC Platinum* se han obtenido los detalles relacionados con la operación de absorción tales como la fecha de anuncio (DateAnnounced) o la fecha de la operación (DateEffective).

Para trabajar con una fuente de datos única hemos decidido unir los datos de ambas fuentes a través de su identificador único común en un panel llamado *DetalleEmpresa*. Del resultado de esta unión, resulta un panel de un total de 524.917 observaciones con 51 atributos ordenados en filas y columnas. El panel nos proporciona información sobre las

#### 4. DATOS

características de una empresa para cada cuatrimestre de un año, desde la primera vez que aparece en *Compustat* hasta que finalmente es adquirida o fusionada o hasta su disolución.

	<b>C<sub>1</sub></b>	<b>C<sub>2</sub></b>	<b>C<sub>3</sub></b>	<b>C<sub>4</sub></b>	<b>C<sub>5</sub></b>	<b>C<sub>n</sub></b>	<b>C<sub>51</sub></b>
<b>F<sub>1</sub></b>	E <sub>1,1</sub>	E <sub>1,2</sub>	E <sub>1,3</sub>	E <sub>1,4</sub>	E <sub>1,5</sub>	E <sub>1,n</sub>	E <sub>1,51</sub>
<b>F<sub>2</sub></b>	E <sub>2,1</sub>	E <sub>2,2</sub>	E <sub>2,3</sub>	E <sub>2,4</sub>	E <sub>2,5</sub>	E <sub>2,n</sub>	E <sub>2,51</sub>
<b>F<sub>3</sub></b>	E <sub>3,1</sub>	E <sub>3,2</sub>	E <sub>3,3</sub>	E <sub>3,4</sub>	E <sub>3,5</sub>	E <sub>3,n</sub>	E <sub>3,51</sub>
<b>F<sub>4</sub></b>	E <sub>4,1</sub>	E <sub>4,2</sub>	E <sub>4,3</sub>	E <sub>4,4</sub>	E <sub>4,5</sub>	E <sub>4,n</sub>	E <sub>4,51</sub>
<b>F<sub>5</sub></b>	E <sub>5,1</sub>	E <sub>5,2</sub>	E <sub>5,3</sub>	E <sub>5,4</sub>	E <sub>5,5</sub>	E <sub>5,n</sub>	E <sub>5,51</sub>
<b>F<sub>n</sub></b>	E <sub>n,1</sub>	E <sub>n,2</sub>	E <sub>n,3</sub>	E <sub>n,4</sub>	E <sub>n,5</sub>	E <sub>n,n</sub>	E <sub>n,51</sub>
<b>F<sub>524.917</sub></b>	E <sub>524.917,1</sub>	E <sub>524.917,2</sub>	E <sub>524.917,3</sub>	E <sub>524.917,4</sub>	E <sub>524.917,5</sub>	E <sub>524.917,n</sub>	E <sub>524.917,51</sub>

*Tabla 2 - Estructura del panel de datos DetalleEmpresa.*

Entre el total de observaciones del panel el número de empresas que compone el estudio es un total de 16.641, americanas, cotizadas en bolsa, correspondientes a diferentes sectores y también con distinto tamaño. Adicionalmente, el panel se puede dividir en dos grupos en función del factor de clasificación que se está considerando, en nuestro caso si la empresa ha sido adquirida o no. El número de filas resultante para cada caso es de 2.768 para las operaciones de fusión o adquisición satisfactorias, que representa un 0.528% del total y 522.149 para las que no lo han sido, un 99.472%.

Se adjunta en el Apéndice I un detalle del panel con la descripción de sus campos [*Apéndice 1: Detalle del panel de datos*].

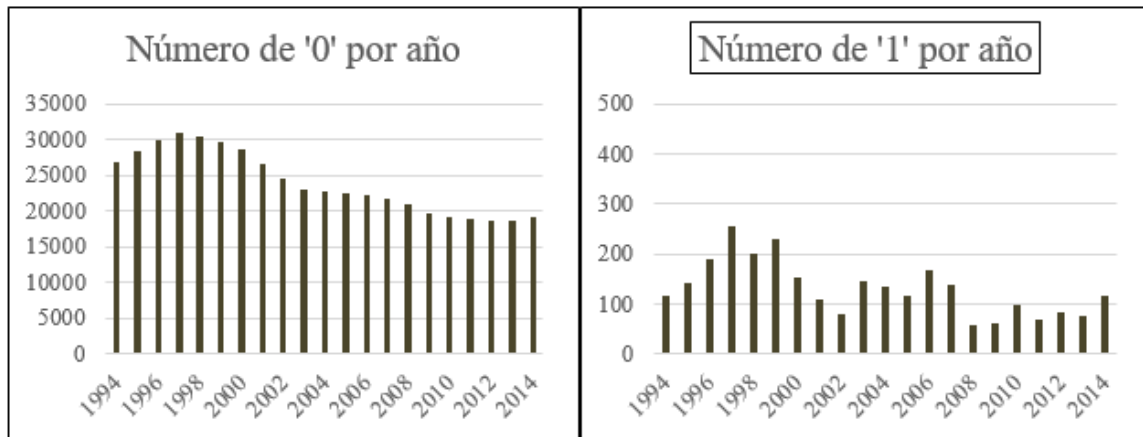


Tabla 3 – Número de operaciones '0' y '1' contempladas en el estudio. En la tabla de la izquierda observamos las operaciones '0' 1994-2004 con una media de 23.500 empresas que no han sido absorbidas. En la tabla de la izquierda observamos las operaciones '1' 1994-2004 con una media de 120 empresas que no han sido absorbidas

#### 4.1.2 Calidad de los datos

Con el objetivo de asegurar la calidad de los datos, en primer lugar se realizará una revisión de las variables que estén sin informar o informadas a nulo, una vez identificadas se observará el impacto de estas variables en el estudio y se procederá a corregirlas. Se identifican dos grupos de variables con esta incidencia, el primero relacionado con detalles de las empresas y el segundo con los detalles de las operaciones. Ambos grupos engloban un número alto de variables que siguen los mismos patrones, por lo cual, se tomarán medidas globales para corregir cada casuística y resolver el mayor número de errores a través del menor número de acciones posible.

#### 4. DATOS

ID	Variables	Problema
GC1	gvkey, lpermno, datadate, fqyearq.fqrt, indfmt consol, popsrc, datafmt, cusip, curedq, datafqrt, atq, ceqq, dlttq, niq, saleq, uniamiq, epspiy, costat, preccq, sic, size, roe, firm, growth, per	Se han identificado 88 filas que no tienen informados ninguno de los campos indicados.
GC2	DateAnnounced, DateEffective, TargetName, AcquirorName, ofSharesAcq, AcquirorName, ofSharesAcq, AcquirorCUSIP, AcquirorPrimarySICCode, ValueOfTransactionmil, TargetCusip, MasterDealType, Form, TargetUltimateParentCUSIP, ofStock, q, AcquirorUltimateParentCUIISP, ofCash, year, month,	Se han identificado 519.371 filas que no tienen informados ninguno de los campos indicados.

*Tabla 4- Incidencias detectadas en el panel inicial de datos. Se dividen en dos grupos GC1 y GC2. El primer grupo contiene las incidencias referentes a filas sin informar relacionadas con la información de la empresa. El segundo grupo contiene las incidencias referentes a filas sin informar relacionadas con la información de la operación de fusión o adquisición.*

En segundo lugar, observamos si existe o no presencia de atípicos para decidir si debemos realizar modificaciones en los datos para que se ajusten a las necesidades del estudio. Se puede observar que las desviaciones estándar en la mayoría de las variables son bastante altas con presencia de valores máximos y mínimos muy alejados de la media. A continuación, se muestran algunos ejemplos.

	Media	Des. St.	Mínimo	Máximo	Observaciones
Atq	7.306	67.453,6	0,00	38.790.000	524.829
Preccq	13,50	851,26	0,00	141.600	524.829
Roe	1,3	959,9	-8653	695.300	524.829

*Tabla 5- Ejemplo sobre la presencia de valores atípicos en variables relevantes del estudio..*

Finalmente, se realiza un análisis en relación a la proporción de datos que existen entre las dos clases identificadas en el panel. El número de observaciones con empresas no adquiridas o fusionadas es significativamente más grande que el número de empresas que sí lo son,  $\approx 191$  veces mayor. Esto es probable que genere un desbalanceo de clases en el momento de la predicción ya que, para aumentar el acierto, el algoritmo puede incurrir en el problema de no establecer diferencia entre clases y tender a clasificar por defecto todas las empresas como no adquiridas.

#### 4.1.3 *Variable dependiente*

Como variable dependiente se ha seleccionado ***merger*** que indica si la empresa ha sido adquirida a final del cuatrimestre correspondiente a su año o no. Esta es una variable dicotómica, toma el valor '1' en el caso de que se haya realizado satisfactoriamente una operación de fusión o una de adquisición y el valor '0' en caso contrario. Se ha seleccionado esta variable ya que es el identificador más representativo de nuestro problema de negocio, si su valor es '1' provoca que aumente el precio de la acción y reduce las posibles fluctuaciones del precio en el corto plazo, lo que resulta más interesante para los inversores.

#### 4.1.4 *Estadísticas descriptivas*

De las 51 variables representadas en el panel para cada observación, las que proporcionan información sobre el estado financiero de la empresa, tienen mayor correlación con la posibilidad de que exista una operación de fusión o adquisición. Las que poseen información relativa a identificadores (empresa, sector) o a fechas (cuatrimestre, año, fecha de realización de la operación) no se puede descartar que no tengan relación, pero sí que es menor que para las anteriores. Por lo tanto, a continuación, presentamos un resumen de las estadísticas descriptivas de las principales variables que influyen en la variable dependiente agrupándolas en función a los valores que esta puede tomar.

#### 4. DATOS

	Merger	Media	Des. St.	Min.	Mediana	Max.	Obs.	Fuente
Atq	1	16.300	117162	4,5	744,4	2.129.000	2.743	Compustat
Atq	0	7.259	67.092	0	338	3.879.000	522.086	Compustat
Ceqq	1	2.370	12.318	-6.898	169	211.700	2.743	Compustat
Ceqq	0	1.311	6.844	-106.400	112,80	255.600	522.086	Compustat
dltqq	1	2.989	-	0	95	413.300	2.743	Compustat
dltqq	0	1.276	-	-45	25	30.480.000	522.086	Compustat
Niq	1	60,440	398,87	-4.139	2,873	5.893	2.743	Compustat
Niq	0	37,25	419,4	-61.660	1,42	69.760	522.086	Compustat
saleq	1	902,90	3.765,3	-3,01	61,24	59.230	2.743	Compustat
saleq	0	638,40	3.303	-25.620	44,36	195.800	522.086	Compustat
Per	1	33,410	171,93	-2.564	23,3	2.672	2.743	Compustat
Per	0	26,09	201,15	-18.630	18	17.600	522.086	Compustat
Roe	1	0,0103	-	-11,95	0,025	10,59	2.743	Compustat
Roe	0	1,3	-	-8.653	0	695.300	522.086	Compustat
leverage	1	0,7967	-	-484,90	0,46	139,7	2.743	Compustat
leverage	0	97	-	-19.210	0	49.610.000	522.086	Compustat

*Tabla 6- Estadísticos descriptivos. Esta tabla proporciona un resumen estadístico para las principales variables independientes que influyen en la variable dependiente. Se encuentran divididas en dos grupos en función de los valores que puede tomar dicha variable.*

- *atq*: cantidad total de los activos que presenta la empresa.
- *ceqq*: capital Social de la empresa.
- *dltqq*: deuda a largo plazo que consiste en préstamos y obligaciones financieras a más de un año.
- *niq*: beneficio después de intereses e impuestos.
- *saleq*: ingresos totales de la compañía, ingresos en efectivo y otros ingresos.
- *per*: ratio precio-beneficio.

- *roe*: ratio de rentabilidad financiera. Relaciona el beneficio económico con los recursos necesarios para obtener ese lucro.
- *leverage*: nivel de apalancamiento de la empresa.

Desde un punto de vista general podemos observar que no existe una homogeneidad clara en los datos, con presencia de valores atípicos que podrían introducir ruido en la predicción. Definiremos ruido como cualquier propiedad de un patrón detectado, que no se debe realmente al modelo subyacente, sino a una aleatoriedad en el conjunto de datos (Duda, Hart, & Stork, 2000). Para solucionar estos problemas, será necesario tomar medidas como la winsorización, normalización y centralización de los datos y de esta forma poder realizar una predicción más fiable. Igualmente, a grandes rasgos si podemos observar que las empresas que son adquiridas cumplen algunos parámetros esperados: cantidad alta de activos en el balance, mayor cantidad de deuda a largo plazo y menor ratio de rentabilidad financiera. Por el contrario, se pueden identificar valores que no son esperados como el nivel de apalancamiento, que resulta ser mayor en empresas no fusionadas o adquiridas que en las que sí lo son.

## 4.2 Preparación de los datos

Tras la realizar la comprensión de los datos y tener una imagen clara y global del tipo y la estructura de información que se dispone, procederemos a preparar los datos para que tengan una estructura y formatos utilizables acordes al problema de clasificación que estamos abordando.

### 4.2.1 Selección de variables

Como hemos identificado en el apartado anterior [4.1.2Calidad de los datos] existen valores que no proporcionan información para la predicción y que por lo tanto sería necesario eliminar para evitar potenciales errores y obtener un panel en el que todos sus

#### 4. DATOS

elementos proporcionan información al modelo. Del mismo modo, debido a que la cantidad de registros con los que estamos trabajando es bastante alta, se ha decidido seleccionar las variables más significativas de las que disponemos para evitar largos tiempos en la computación del resultado final del modelo.

En primer lugar, se tratan las variables que estén sin informar o informadas a nulo [Tabla 4]. Para el *GCI* se decide eliminar los 88 registros que no tenían informados los campos relativos a los estados financieros de las empresas. Al tratarse de un número bajo de campos sin informar se realizará una eliminación de las filas que presentan esta incidencia. Se eliminarán también los 519.371 valores que no presentan información sobre los detalles de la operación absorción *GC2*. En este caso se ha decidido eliminar las columnas ya que el número de casos es considerablemente alto. Como resultado, las operaciones de absorción exitosas pasarán a ser un total de 2.743 y las no exitosas 522.086, reduciéndose en 25 y 63 observaciones respectivamente y el número de columnas se reduce en 20. El conjunto de datos resultante lo formará un panel compuesto por 524.829 filas y 31 columnas.

En segundo lugar, se realizará un proceso de selección de variables que tiene como objetivo reducir la complejidad computacional del problema para grandes cantidades de datos. Se ha observado que existen algunas variables que debido al poco valor que aportan para la clasificación no resulta óptimo introducirlas en la predicción. Al tratarse de un número significativamente alto de observaciones que no aportan información al modelo, no excluirlas supondría aumentar el coste computacional de la predicción y aumentar la probabilidad de error al incluir datos que no influyen en la probabilidad de que exista una operación de fusión o adquisición. También, habría que asumir los cambios de formato de estas variables, que implica alto coste de programación por una mejora en los resultados casi inapreciable.



La decisión de eliminar estas variables se ha tomado en base a la versión estadística del principio conocido como *Navaja de Occam* (Di Deco Sampedro & Díaz García, 2012). Debido a que en el dominio de operaciones de fusión y adquisición es muy costoso recoger información valiosa de las empresas, es preferible un sistema automático que sea capaz de ofrecer un buen comportamiento con menos información a otros que requieran más datos para conseguir un comportamiento similar, entendiendo menos información como menor número de atributos, menor número de observaciones o ambas.

El panel está compuesto por treinta y una variables de las cuales: diecisiete se corresponden a fechas (*qdate*, *datadate*, *datafqtr*), clasificadores (*indfmt*, *consol*) e identificadores (*gvkey*, *cusip*); doce a variables que proporcionan información sobre los estados financieros de la empresa (*atq*, *sale*, *niq*) y una que nos indica si la empresa ha sido adquirida o fusionada o no (*merger*). Para la configuración del panel final se han seleccionado ocho variables consideradas las más representativas entre el conjunto total de variables para predecir una absorción.

De acuerdo con (Meador et al., 1996; Palepu, 1986) existen variables de carácter financiero que pueden ser utilizadas para la formulación de hipótesis de probabilidad de adquisición. Estas variables proporcionan información sobre el estado financiero de la empresa y son las siguientes: *atq*, *ceqq*, *dlttq*, *niq*, *saleq*, *per*, *roe*, *leverage*. A partir de este conjunto de ocho variables de las que disponemos más de 500.000 observaciones intentaremos predecir el valor de la variable dependiente *merger*.

Con estas variables conseguiremos verificar al menos cuatro de las seis hipótesis propuestas por (Palepu, 1986) y cuatro de las diez propuestas por (Meador et al., 1996) que también incluye en su estudio variables relacionadas con el mercado.

#### 4. DATOS

- *Profitability* (Meador et al., 1996) o *Inneficient management hypothesis* (Palepu, 1986). **Rentabilidad**, aparece descrita por la variable *roe* que relaciona el beneficio neto como un porcentaje de la aportación de capital de los accionistas. Mide la rentabilidad de una empresa relacionando el beneficio obtenido con la aportación de los accionistas, cuanto mayor sea el *roe*, mayor rentabilidad tiene la empresa. Según (Palepu, 1986) las empresas que presentan mala rentabilidad tienen más probabilidad de convertirse en objetivo de adquisición.
- *Size* (Meador et al., 1996) o *Size hypotesis* (Palepu, 1986). **Tamaño**, se mide a través de las variables *saleq*, *atq*, *niq*. Que indican la cantidad de ingresos, el número de activos y el beneficio después de impuesto respectivamente. Cuanto mayores sean estas variables mayor es el tamaño de la empresa y según (Palepu, 1986) a mayor tamaño menor probabilidades de compra.
- *Leverage* (Meador et al., 1996). **Apalancamiento**, nos lo indica la variable leverage que mide cuanto capital es financiado por deuda y evalúa la capacidad de una empresa para cumplir obligaciones financieras. Este apartado también se ve completado por la variable *dltqq* que es la cantidad de deuda a l/p. (Meador et al., 1996) a menor apalancamiento mayor probabilidad de adquisición.
- *Price-earnings* (Meador et al., 1996) o *price-earnings hypothesis* (Palepu, 1986). Ratio **precio-beneficio**, es la relación entre el precio y los beneficios su valor indica cuántas veces se está pagando el beneficio neto anual de una empresa determinada al comprar una acción de ésta. (Palepu, 1986) afirma que las empresas con un *per* bajo es más probable que sean compradas.

#### 4.2.2 *Formateo de los datos*

El conjunto de datos con el que se realiza el experimento está formado por valores continuos y no es necesario realizar cambios de formato en las variables independientes, debido a que los algoritmos de clasificación son capaces de trabajar con este tipo de datos. Por el contrario, la variable dependiente al ser una variable dicotómica que toma valores numéricos si necesita ser expresada como una variable discreta. Para ello, se le realizará un cambio de formato a una variable tipo *factor* considerada en este experimento como una variable que solo puede tomar valores dentro de un conjunto numerable (0,1).

#### 4.2.3 *Winsorización y normalización*

Tras la selección y el formateo de las variables indicadas se procederá a realizar una homogenización de los datos, para principalmente de esta forma minimizar el ruido que puedan generar los valores atípicos [Tabla 5]. Este ajuste de los datos se realizará a través de dos pasos el primero de winsorización y el segundo de normalización.

El proceso de winsorización se realizará de forma univariante y tendrá un nivel del 1% en las variables: *atq*, *ceqq*, *dlttq*, *niq*, *saleq*, *per*, *roe*, *leverage*. Sustituiremos los valores del primer percentil por el primer valor del segundo y del último percentil por el último valor del antepenúltimo.

El segundo paso es la normalización que consiste en centrar y escalar los datos, de este modo permitiremos comparaciones independientes de la unidad de medida. Centrar la variable consiste en sustraer su media a cada uno de sus valores iniciales y escalar los datos consiste en dividir todos sus valores por su desviación típica. Al realizar el proceso de normalización obtendremos una media nula y una desviación típica igual a 1. Finalmente, lo que obtenemos con esto son datos independientes de la unidad y variables que poseen la misma dispersión y la misma media.

#### 4. DATOS

## Capítulo 5

### Estudio empírico.

En este apartado se describirán los recursos utilizados desde la obtención del conjunto de datos final hasta la implantación de los modelos de clasificación y de evaluación de resultados. La sección consta de una introducción teórica sobre el problema de clasificación y las técnicas, software y hardware utilizado y concluye con la presentación de los modelos adaptados para el problema a resolver.

#### 5.1 Problema de clasificación

Para la predicción de absorción de empresas realizada en este estudio vamos a emplear técnicas de *clasificación supervisada*, también es conocida como *reconocimiento de patrones* o *discriminación estadística*. El término de *supervisada* se refiere al hecho de que el conjunto de datos utilizado para la creación de la regla de clasificación está supuestamente bien clasificado, típicamente usando procedimientos no estadísticos externos. De este modo proporcionan una base fiable para la clasificación de una nueva observación (Baíllo & Cuevas, 2008).

El problema de clasificación se basa en decidir o predecir dadas  $M+1$  poblaciones o clases (e.j. empresas absorbidas o empresas no absorbidas) a cuál de las clases pertenece una nueva observación cuya clase desconocemos. Una *observación* es una colección de mediciones numéricas y formalmente se define como un vector  $\mathbf{x}$   $d$ -dimensional. La *clase* se denota como  $y$  que toma valores en un conjunto finito  $\{0, 1, 2, \dots, M\}$  (para nuestro caso

## 5. EXPERIMENTACIÓN

0 y 1). El objetivo de la clasificación es definir una función  $g(x): \mathbf{R}^d \rightarrow \{0, \dots, M\}$  que asigne una clase  $y$  a un nuevo dato  $x$ . La aplicación  $g$  recibe el nombre de *clasificador*. Un clasificador fallará para el par  $(x, y)$  si  $g(x) \neq y$  (Devroye, Györfi, & Lugosi, 2014).

Como hemos comentado en el párrafo anterior la *clasificación supervisada* se encarga de decidir si una observación de una variable aleatoria  $\mathbf{X}$  (con valores en un *espacio de características*  $F$  dotado de dimensión  $d$ ), pertenece a la población  $\mathbf{P}_0$  o  $\mathbf{P}_1$ . La decisión debe ser tomada con la información proporcionada por el *conjunto de entrenamiento* (observaciones bien clasificadas)  $\mathbf{D}_n = \{(\mathbf{X}_i, \mathbf{Y}_i), 1 \leq i \leq n\}$  donde  $\mathbf{D}_n$  está formado por  $n$  observaciones de pares  $(\mathbf{X}_i, \mathbf{Y}_i)$  independientes. Por lo tanto, el problema matemático que se pretende resolver es la búsqueda de un *clasificador*  $g_n(x) = g_n(x; \mathbf{D}_n)$  con  $g_n : F \rightarrow \{0, 1\}$ , que minimice el error de clasificación  $\mathbf{P}(g_n(\mathbf{X}) \neq \mathbf{Y})$  (Baíllo & Cuevas, 2008).

En nuestro caso definimos  $\mathbf{X}$  como el vector que contiene el conjunto de variables (*atq*, *ceqq*, *dlttq*, *niq*, *saleq*, *per*, *roe*, *leverage*) correspondientes a las características de las empresas y la variable  $\mathbf{Y}$  representará la variable dependiente *merger*. Trataremos de clasificar las observaciones en dos clases  $\mathbf{P}_0$  para las empresas en la que no se realiza la operación y  $\mathbf{P}_1$  en las que sí. El *conjunto de entrenamiento*  $\mathbf{D}_n$  está formado por una selección de  $n$  observaciones independientes de  $\mathbf{X}$  pertenecientes al panel final de datos y las etiquetas  $\mathbf{Y}$  correspondientes. En nuestro estudio estos datos corresponderían al panel de observaciones *DetalleEmpresa* que conforma un histórico de operaciones de fusiones o adquisiciones entre empresas.

Como podemos observar, la probabilidad de clasificar correctamente una observación  $x$  está relacionada con el error que posea el *clasificador*  $L=L(g)$ . El objetivo de la clasificación será minimizar este error y de este modo conseguir la menor probabilidad de incurrir en él. Para ello lo ideal sería encontrar el clasificador óptimo  $g^*(x) = \mathbf{1}_{\{\eta(x) > 1/2\}}$  donde  $\eta(x) = E(Y|X=x)$ , que depende de la distribución de  $(\mathbf{X}, \mathbf{Y})$ . En caso de que esta

distribución sea conocida, el clasificador óptimo puede ser calculado, pero la mayoría de las veces la distribución es desconocida y por lo tanto  $\mathbf{g}^*$  no se puede calcular.

El proceso de construir un  $\mathbf{g}_n$  que minimice el error de clasificación se llama *entrenamiento*. Se pueden definir distintos *clasificadores* con distintas propiedades en función del conjunto de datos  $(\mathbf{X}_n, \mathbf{Y}_n)$  del que disponemos. Para este estudio consideraremos las siguientes reglas de clasificación: el *discriminante lineal* que consiste en encontrar una dirección óptima que separe las observaciones entre dos clases. *K-vecinos cercanos* que clasifica una observación en función de las clases de las observaciones más cercanas a ella. Las *máquinas de vector soporte*, que a través de la técnica *kernel trick* son capaces de clasificar observaciones que no tienen separación lineal en el plano. *Random Forest* que consiste en una combinación de árboles estructurados. Se ha configurado esta elección de clasificadores que contempla dos técnicas simples y tradicionales como LDA y KNN y otras dos más complejas, actuales en el estado del arte como SVM y RF. Sin embargo, otros ejemplos de clasificadores que pueden encontrarse en manuales de referencia Bishop (2006) Duda et al. (2000) podrían tener cabida en el estudio.

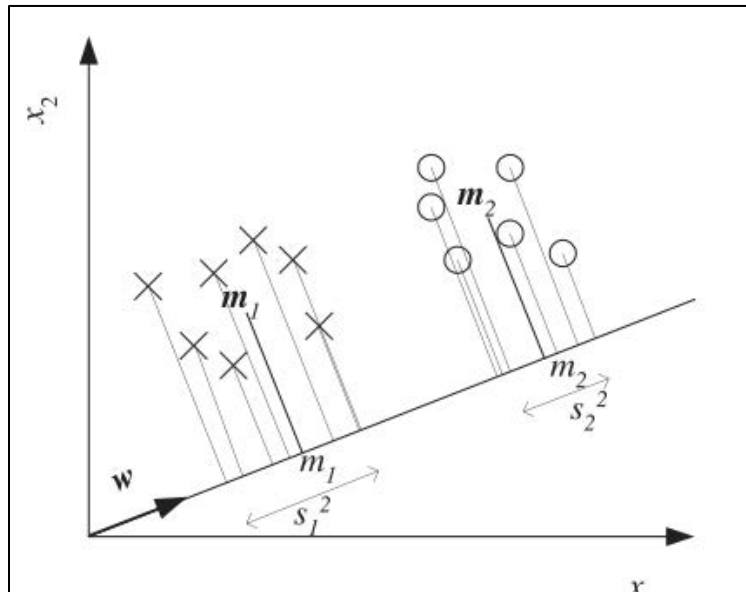
## 5.2 Técnicas de clasificación

### 5.2.1 Análisis discriminante lineal

El Análisis Discriminante Lineal o Linear Discriminant Analysis (LDA), también conocido como discriminante de Fisher, busca la mejor separación lineal de las clases. Consiste en la reducción de las dimensiones del problema a través de la combinación lineal de rasgos que caracterizan o separan las poblaciones maximizando la distancia entre las medias de las clases proyectadas y controlando la varianza de cada clase (Duda et al., 2000).

## 5. EXPERIMENTACIÓN

Definiendo un problema de clasificación entre dos clases  $P_0$  y  $P_1$  se quiere encontrar la dirección, definida como un vector  $w$ , tal que cuando los datos son proyectados sobre  $w$ , las observaciones de ambas clases se encuentran lo más alejadas posible.  $z=w^t x$  es la proyección de  $x$  sobre  $w$  y por lo tanto una reducción de la dimensionalidad  $d$  a 1. Finalmente el dato se asigna a la clase cuya media es más cercano.



*Ilustración 4 – Ejemplo de un Discriminante Linear para dos clases. Se busca una dirección  $w$ , tal que la distancia entre la media de las dos clases proyectadas hacia esa dirección ( $m_1, m_2$ ) es maximizada y la varianza sobre las medias  $s_1$  y  $s_2$  es minimizada controlando la varianza de cada clase (Bishop, 2006).*

Este proceso se ilustra en la [Ilustración 4] donde  $m_1$  y  $m_2 \in \mathbb{R}^d$  son las medias de las observaciones antes de la proyección y  $m_1$  y  $m_2 \in \mathbb{R}$  después de la proyección (de acuerdo a la notación utilizada la clase  $m_2$  sería  $m_0$ ). Después de la proyección, para que las dos clases estén correctamente separadas, queremos que las medias de ambas observaciones estén lo más alejadas y agrupadas posibles. Para ello es necesario aumentar la distancia entre  $m_1$  y  $m_2$  y reducir la suma de  $s_1^2 + s_2^2$  (siendo  $s_k$  la varianza de los elementos de la

40



clase  $k$  tras la proyección). El discriminante lineal de Fisher es aquel con el cual la  $\mathbf{w}$  **maximiza**

$$J(\mathbf{w}) = \frac{(\mathbf{m}_1 - \mathbf{m}_2)^2}{s_1^2 + s_2^2}. \quad (1)$$

A partir de la definición del discriminante lineal de Fisher, se puede llegar a la expresión para el cálculo de la  $\mathbf{w}$  que maximiza la distancia entre clases y minimiza la distancia entre las observaciones de una clase, donde  $s_w$  es la matriz de covarianzas de la clase.

$$\mathbf{w} = s_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \quad (2)$$

### 5.2.2 $k$ - vecinos cercanos

La técnica de  $k$ -vecinos cercanos es una extensión de la regla de estimación por vecinos cercanos, desarrollado por Fix & Hodges (1951). La clasificación de una observación  $\mathbf{x}$  se basa en la *clase* a la que pertenecen las  $k$  observaciones más cercanas a la observación a clasificar  $\mathbf{x}$ . Normalmente, la proximidad se mide a través de la distancia Euclídea entre los diferentes los datos (Cherkassky & Mulier, 2007).

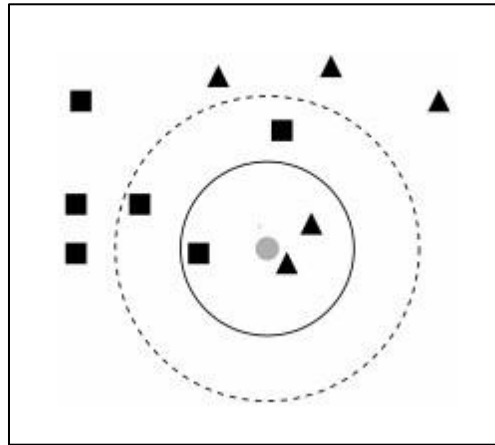
$$\mathbf{g}_n(\mathbf{x}) = \begin{cases} 1, & \frac{1}{k} \sum_{i=1}^n y_i K_k(\mathbf{x}, \mathbf{x}_i) > 0.5 \\ 0, & \text{en el resto de los casos} \end{cases}. \quad (3)$$

donde  $K_k(\mathbf{x}, \mathbf{x}_i) = 1$  si  $\mathbf{x}_i$  es una de las  $k$  observaciones cercanas a la observación a clasificar y 0 en caso contrario. El conjunto de funciones aproximadas es  $\mathbf{g}_n(\mathbf{x}) = \mathbf{y}$ , donde  $\mathbf{y}$  puede tomar los valores discretos  $\{0, 1\}$ . El riesgo empírico asociado a este clasificador es

## 5. EXPERIMENTACIÓN

$$L_{\text{emp\_local}}(\mathbf{g}_n(\mathbf{x})) = \frac{1}{k} \sum_{i=1}^n (Y_i - y)^2 K_k(\mathbf{x}, \mathbf{x}_i), \quad (4)$$

El riesgo empírico de una mala clasificación es minimizado cuando  $y$  adopta el valor de la mayoría de las etiquetas de clase cercanas a la observación.



*Ilustración 5 – k- vecinos cercanos. La función de clasificación para los k-vecinos cercanos empieza en el círculo gris y crece de forma esférica hasta que engloba una cantidad de observaciones de tamaño k, posteriormente clasifica el círculo gris dependiendo de la mayoría de votos de los k vecinos más cercanos. En este caso con un  $k=3$ , el círculo pasaría a clasificarse como un triángulo y para un  $k=5$ , pasaría a clasificarse como un cuadrado. (Fuente: [https://es.wikipedia.org/wiki/K-vecinos\\_m%C3%A1s\\_cercanos#/media/File:KnnClassification.svg](https://es.wikipedia.org/wiki/K-vecinos_m%C3%A1s_cercanos#/media/File:KnnClassification.svg))*

En la práctica hay que estimar  $k$  [Ilustración 5] (el número de vecinos cercanos) ya que juega el papel de un *parámetro de suavizado*. En general  $k$  debe ir creciendo cuando aumente  $n$  aunque de forma gradualmente más lenta (Hastie, Tibshirani, & Friedman, 2001).

### 5.2.3 Máquinas de vectores soporte

Las máquinas de vector soporte o Support Vector Machines (SVM) son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik. SVM es un sistema de clasificación muy potente ya que son capaces de entrenar de manera eficiente clasificadores lineales en un ‘espacio kernel inducido’ respetando los conocimientos proporcionados por la teoría de la generalización y aprovechando la teoría de la optimización (Cristianini & Shawe-Taylor, 2000). El funcionamiento de las SVM consiste en la construcción de un hiperplano de separación óptimo en un espacio de dimensionalidad muy alta (o incluso infinita) que sea capaz de separar los datos en las clases que deseamos. Una buena separación entre las clases permitirá una clasificación correcta de la nueva muestra. Para esto será necesario maximizar la distancia (margen) del hiperplano a los puntos más cercanos de las dos clases vectores soporte.

El *espacio kernel inducido* o se construye a partir de un mapeado no lineal  $\phi()$  que permite proyectar los datos a un espacio de dimensionalidad mayor para que estos puedan ser separados por el hiperplano  $\mathbf{H}$  más fácilmente. Asumimos que cada atributo de la observación  $\mathbf{x}_k$  se transforma en  $\mathbf{t}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \{ \phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_n) \}$  para  $k=1,2,3,\dots,n$ , a través de una transformación lineal posicionando la observación a un lado o al otro del hiperplano dependiendo de la clase a la que pertenezcan las observaciones  $\mathbf{P}_0$  o  $\mathbf{P}_1$ . Obteniendo de este modo un discriminante lineal en un espacio aumentado (Duda et al., 2000).

## 5. EXPERIMENTACIÓN

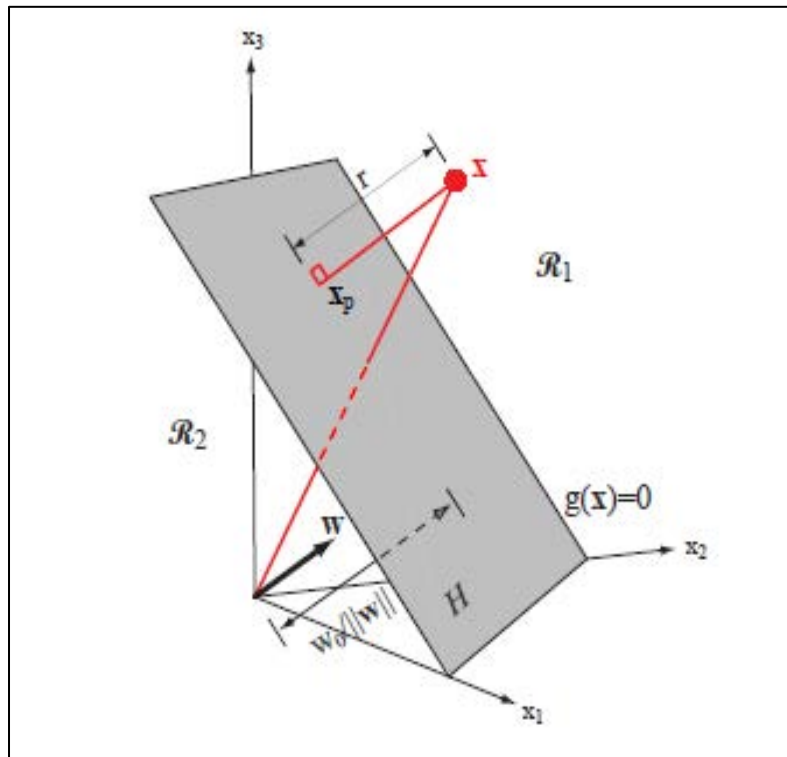
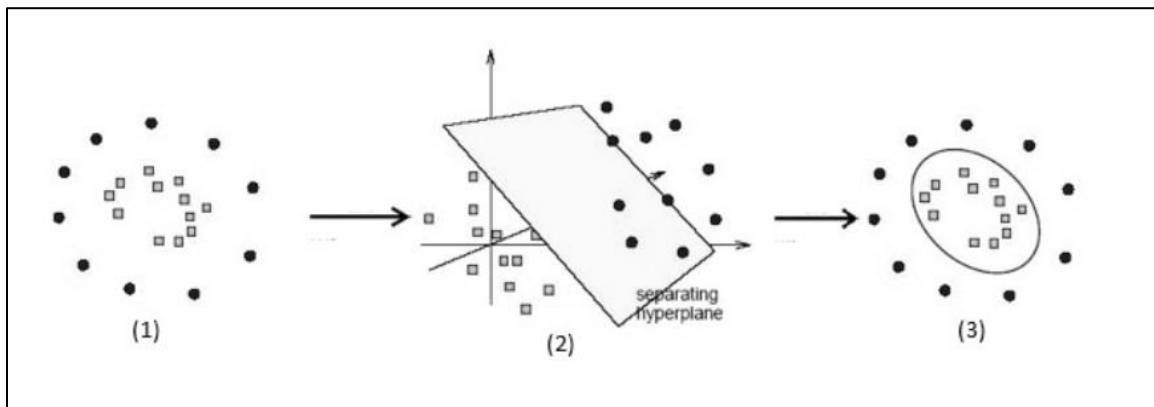


Ilustración 6 – La frontera de decisión liner  $H$ , donde  $g(x)=w'x+w_0$ , separa las observaciones del espacio en dos sub-espacios  $\mathcal{R}_1$  (donde  $g(x)>0$ ) y  $\mathcal{R}_2$  (donde  $g(x)<0$ ) (Duda et al., 2000)

En resumen, el objetivo de las máquinas de vector soporte es encontrar el hiperplano separador que disponga de un mayor margen, ya que esperamos que cuanto mayor sea el margen, mejor será la generalización del clasificador (Duda et al., 2000). La potencia de SVM radica en que para hacer todo esto no es necesario conocer  $\mathbf{H}$  ni  $\phi$ , nos basta con calcular los productos escalares de las imágenes mediante funciones kernel. Esto es lo que se conoce como *kernel trick* y la elección de distintos kernel influirá en las propiedades del SVM (Duda et al., 2000).



*Ilustración 7 – Ejemplo de problema de clasificación que no es linealmente separable. No es posible resolverlo linealmente en el plano y por lo tanto es necesario proyectarlo a un espacio de dimensiones mayores. (1) Espacio original donde se encuentran representados los datos. (2) Se realiza la transformación y se obtiene  $y_k = \phi(x_k)$ , un espacio de dimensiones mayores. En este espacio de características el objetivo es encontrar el hiperplano óptimo, aquel que maximice la distancia con las observaciones más cercanas. Los vectores soporte son aquellas observaciones más cercanas, las que se encuentran a distancia  $b$  del hiperplano. (3) Representa el espacio original con la separación en clases realizada (Duda et al., 2000).*

#### 5.2.4 Random Forest

Random Forest es un clasificador que consiste en una combinación de árboles de decisión  $\{g(x, D_n^k), k=1, 2, \dots\}$  generados a partir de distintos conjuntos de entrenamiento. Cada árbol realiza un único voto a favor de la clase con mayor probabilidad para la observación  $x$  (Breiman, 2001). Como característica de random forest es necesario mencionar que es un método que emplea bagging para combinar diferentes árboles, donde cada uno es construido con variables aleatorias.

En primer lugar, random forest elige subconjuntos de datos al azar, con reemplazo. Segundo, se crea un árbol de decisión para cada subconjunto eligiendo de forma aleatoria

## 5. EXPERIMENTACIÓN

un atributo para realizar la partición en cada nodo. Por último, se asigna a  $x$  la clase mayoritaria entre las predicciones individuales de los árboles.

### 5.3 Recursos utilizados

#### 5.3.1 Equipo

Para la elaboración de este estudio se ha utilizado un equipo HP Pavillion dv6 con un procesador Intel® Core™ i7 – 2670QM CPU @2.20GHz, memoria RAM 8,00GB y un sistema operativo Windows 10 de 64 bits.

#### 5.3.2 R como lenguaje de programación

Se ha seleccionado R como lenguaje de programación ya que es un lenguaje por procedimientos orientado al análisis estadístico, aunque su uso también está presente en otros campos como el cálculo numérico debido a su precisión. Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka del departamento de Estadística de la universidad de Auckland en 1993 (Ihaka, 1998), aunque sus orígenes parten del lenguaje de programación S, desarrollado en los Bell Laboratories de AT&T. Actualmente, R es un proyecto GNU de software libre llevado a cabo por el R Development Core Team y distribuido de forma gratuita.

En áreas como, investigación biomédica, bioinformática estadística y matemáticas financieras R es ampliamente utilizado. Esto se debe entre otras cosas a su programación multiparadigma que mezcla diferentes paradigmas de programación como pueden ser la programación orientada a objetos, por procedimientos o reflexiva. Otra característica de este lenguaje es la robustez que le aportan las librerías desarrolladas que con facilidad se implementan al software.

Los experimentos se desarrollarán utilizando los siguientes paquetes de R.

- Paquete *Caret*: paquete desarrollado por Max Kuhn. Contributions from JedWing, SteveWeston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan, and Tyler Hunt (Max Kuhn Contributions from Jed Wing et al., 2017) en su version 6.-76. **Será utilizado como paquete de referencia con el objetivo de unificar y homogeneizar el código lo máximo posible.** El paquete contiene un misceláneo de funciones para entrenamiento y dibujo de modelos de clasificación y regresión. A partir de este paquete es posible que se realice la importación de paquetes adicionales para la completitud de las operaciones, al no trabajar directamente con estos paquetes se ha omitido su descripción, salvo el paquete *class*, *e107*, *randomForest* que se detallan a continuación.
- Paquete *class*: paquete desarrollado por Brian Ripley y Willian Venables (B. Ripley & Venables, 2015) en su versión 7.3-14. Incluye varias funciones de clasificación de las cuales para el proyecto se utilizará en relación a *k-nearest neighbour*.
- Paquete *e1071*: paquete desarrollado por David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, Friedrich Leisch, (Meyer et al., 2017) en su versión 1.6-8. Incluye un misceláneo de funciones: “*Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier*”(Meyer et al., 2017). De entre las funciones que componen este paquete se utilizarán las relacionadas con máquinas de soporte vectorial.

## 5. EXPERIMENTACIÓN

- Paquete *Random Forest*: paquete desarrollado por Leo Breiman y Adele Cutler para Fortram y Andy Law y Matthew Weiner para R (Horning et al., 2014) en su versión 4.6-12. El uso del paquete es orientado a clasificación y regresión basada en random forest utilizando inputs aleatorios.
- Paquete *MASS*: paquete desarrollado por Brian Ripley, Bill Venables, Douglas M. Bates, Kurt Hornik, Albrecht Gebhardt, David Kuhn (W. N. V. and B. D. Ripley, 2017) en su versión 7.3-47. La paquete se desarrolló con el objetivo de crear un conjunto de funciones y datasets para complementar otros paquetes de los mismos autores. Para este proyecto se utilizará en la implementación del análisis de discriminante linear.

Para desarrollar el código hemos utilizado R Studio que es un entorno de desarrollo integrado (IDE) para R. Incluye una consola, editor de resaltado de sintaxis que soporta la ejecución de código directamente, así como herramientas para el trazado, la historia, la depuración y gestión del espacio de trabajo. Tiene la misión de proporcionar el entorno de programación tanto de informática como de estadística para R y permite un entorno amigable dónde cualquier usuario pueda analizar los datos que desee.

### 5.3.3 *Microsoft Excel*

Excel es un programa informático desarrollado y distribuido por Microsoft Corp. Se trata de un software que permite realizar tareas contables y financieras gracias a sus funciones, desarrolladas específicamente para trabajar con hojas de cálculo (Pérez Porto & Gardey, 2009).

En este experimento utilizaremos Excel para procesar los resultados obtenidos de la predicción de una manera más cómoda y visual gracias a sus hojas de cálculo. Se utilizará



en el uso de técnicas estadísticas básicas y procesos simples para mostrar de una forma más amigable y adecuada para el problema de negocio los resultados obtenidos.

## **5.4 Modelo de clasificación**

### *5.4.1 Conjunto de entrenamiento y de test*

Para poder definir una función que prediga correctamente el mayor número de empresas que serán adquiridas o fusionadas, es necesaria la creación de dos conjuntos de datos. Se crearán simultáneamente, pero tendrán diferentes valores: entrenamiento y test. Esto nos permitirá tener una evaluación más realista evitando posibles problemas de sobreajuste.

El conjunto de datos de entrenamiento servirá para poder entrenar el algoritmo utilizado para predecir la hipótesis deseada. De este modo el algoritmo aprenderá de estos datos (valores, características, relaciones, variaciones...) para posteriormente estar ‘adaptado’ a nuevos casos que serán similares. El conjunto de entrenamiento está configurado por una proporción de empresas adquiridas y no adquiridas 1:1. Como datos de entrenamiento etiquetados como ‘1’ se seleccionan 2/3 del total de observaciones clasificadas como ‘1’. Una vez establecido el número de observaciones para las operaciones satisfactorias se seleccionan la misma cantidad de ‘0’.

Se ha determinado una proporción balanceada para el conjunto de entrenamiento como una medida particular para este problema ya que el número de ‘0’ en el total del panel es claramente superior al de ‘1’. Con esta distribución se intenta evitar que el método de

## 5. EXPERIMENTACIÓN

clasificación tienda a clasificar siempre como '0' todas las observaciones tratando de minimizar el error.<sup>1</sup>

El conjunto de datos de test servirá para poder medir el rendimiento de la t planteada en cada algoritmo utilizado, por lo que deben ser independientes. Esto se consigue introduciendo los datos de test una vez entrenado el algoritmo. Los datos de test son una representación fiel del conjunto de datos de trabajo y por lo tanto tienen mismas proporciones que existen en el total del panel. La relación de empresas adquiridas a no adquiridas ha de ser 1:190. Para realizar la predicción con estos datos se introducen las variables predictoras de cada observación en el clasificador ya entrenado y este realiza la predicción clasificándolas entre adquiridas y no adquiridas, posteriormente se compara este resultado con el valor conocido para cada caso y se analiza el rendimiento del algoritmo.

Para cada caso, nombrando a las empresas que son adquiridas o fusionadas como '1' y a las que no lo son como '0' se configura una división del panel en los siguientes grupos.

	Número de observaciones	% sobre el total
Panel	524.829	100%
1	2.743	0,5%
0	522.086	99,5%

*Tabla 7- Número de observaciones en función del factor merger. La tabla representa el número total de observaciones dividiéndolo en dos casos. '1' Para operaciones con éxito '0' para operaciones sin éxito.*

---

<sup>1</sup> Las predicciones de experimentos previos con diferentes distribuciones en los conjuntos de entrenamiento presentaron los siguientes resultados. Balanceo del conjunto entrenamiento 1:190 un '1' por 190 '0' para knn, svmLin, svmLog, rf (acierto  $\approx$  100%, sensibilidad  $\approx$  0%, precisión  $\approx$  0%) a partir de un conjunto de entrenamiento con este balanceo predice todos los datos como '0'. Balanceo del conjunto de entrenamiento 1:4 un '1' por 5 '0' para knn, svmLin, svmLog, rf (acierto  $\approx$  80%, sensibilidad  $\approx$  0%, precisión  $\approx$  0%) a partir de un conjunto de entrenamiento con este balanceo predice como mínimo el 80% de los datos como '0'.

	Número de observaciones	% sobre el total
Conjunto de entrenamiento total	3.656	100%
‘1’ en conjunto de entrenamiento	1.828	50%
‘0’ en conjunto de entrenamiento	1.828	50%
Conjunto de test total	174.574	100%
‘1’ en conjunto test	914	0,5%
‘0’ en conjunto test	193.660	99,5%

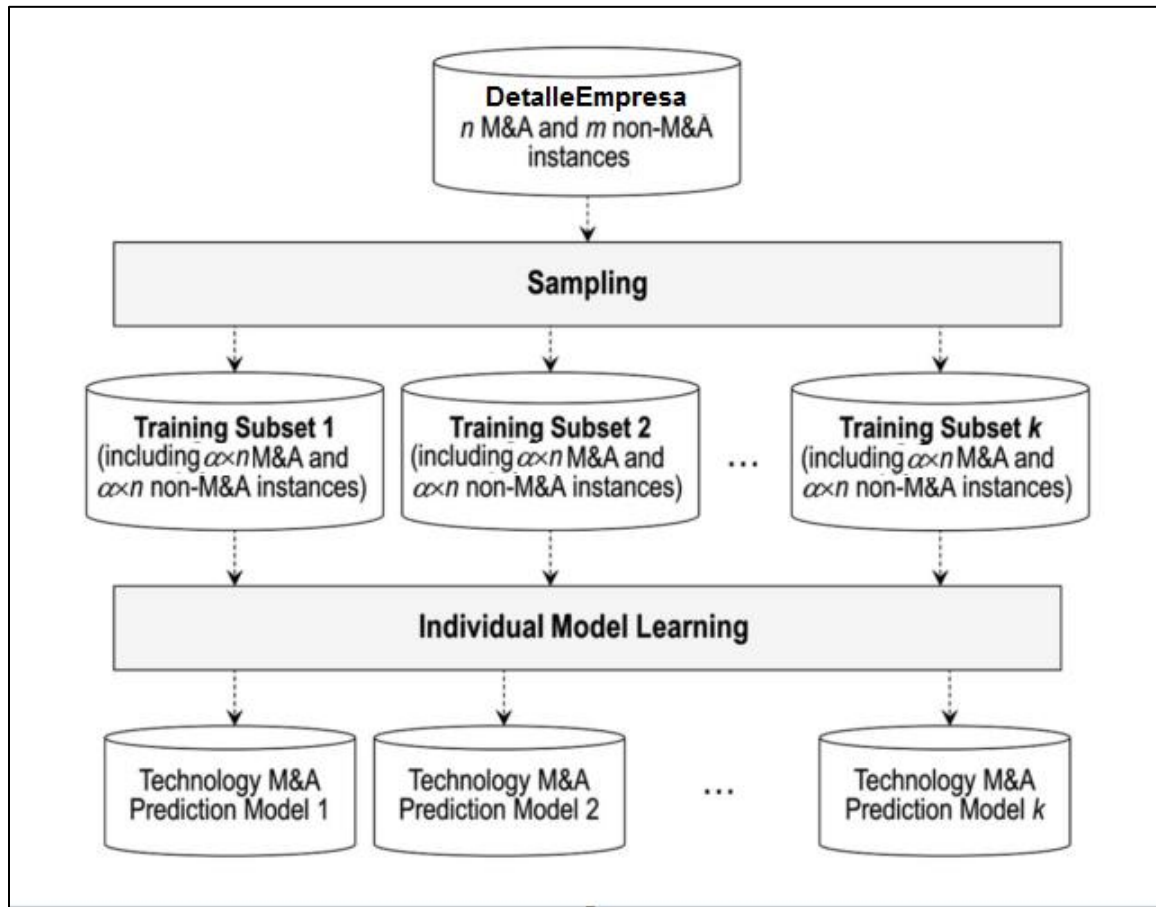
*Tabla 8 – Conjuntos de entrenamiento y de test. La tabla representa los dos conjuntos creados para el entrenamiento y el testeo del clasificador. Se indica una división de los conjuntos en función de los valores que puede tomar la variable dependiente merger ‘1’ y ‘0’.*

Finalmente observamos que existen 3.656 casos con los que se realizará el entrenamiento de los algoritmos, con la misma proporción entre operaciones satisfactorias y no satisfactorias y 174.574 con los que se realizará la clasificación con una proporción de 1 operación satisfactoria por 190 no satisfactorias.

#### 5.4.1 Segmentación

Para intentar minimizar el error global que produce no detectar los ‘1’ hemos decidido decantarnos por la técnica de segmentación (Hastie et al., 2001) [*¡Error! No se encuentra el origen de la referencia.*]. Esta técnica consiste en elaborar diferentes conjuntos de entrenamiento y de test de y con estos conjuntos, de manera iterativa se entrenará el algoritmo y se realizará la predicción. Con el objetivo de obtener una clasificación consistente se repetirá todo el proceso 100 veces con diferentes conjuntos de entrenamiento y test para cada algoritmo.

## 5. EXPERIMENTACIÓN



*Ilustración 8 - Ilustración 9 - Asumiendo que tenemos  $n$  M&A observaciones y  $m$  no-M&A observaciones (donde  $n < m$ ) en el conjunto de datos total. Cada conjunto de entrenamiento dispondrá de  $a \times n$  observaciones de M&A y  $a \times n$  observaciones no M&A, donde  $a = 2/3$  y se usa para ajustar el ratio entre empresas adquiridas y no adquiridas. Posteriormente, un algoritmo de clasificación supervisada aplica entrenamiento para cada conjunto de datos de entrenamiento. De este modo obtendremos  $k$  versiones de un algoritmo para los diferentes conjuntos de entrenamiento y test.*

En el [Apéndice 2: Flujo del proceso] y [Apéndice 3: Aplicación de las técnicas de clasificación] se muestra información relativa al flujo de ejecución del proceso y a la implementación de las técnicas de clasificación.

## 5.5 Modelo de evaluación

Para medir el rendimiento de cada algoritmo se utilizarán las siguientes herramientas.

- Matriz de confusión

		Real	
		0	1
Predicción	0	Verdadero Negativo	Falso Negativo
	1	Falso Positivo	Verdadero Positivo

*Tabla 9 - Matriz de confusión. Se utilizará para evaluar el resultado de la clasificación, divide los resultados en cuatro categorías: Verdadero Negativo, Falso Negativo, Falso Positivo, Verdadero Positivo.*

Las entradas en la matriz de confusión tienen el siguiente significado:

- Verdadero negativo (VN): empresas clasificadas como no adquiridas o fusionadas y que efectivamente no lo están.
  - Falso negativo (FN): empresas clasificadas como no adquiridas o fusionadas, que por el contrario si lo están.
  - Falso positivo (FP): empresas clasificadas como adquiridas o fusionadas, que por el contrario si lo están.
  - Verdadero positivo (VP): empresas clasificadas como adquiridas o fusionadas y que efectivamente lo están.
- Acierto: proporción de resultados que son clasificados correctamente entre el número total de casos.

$$Acierto = \frac{VP + VN}{Total} \quad (5)$$

## 5. EXPERIMENTACIÓN

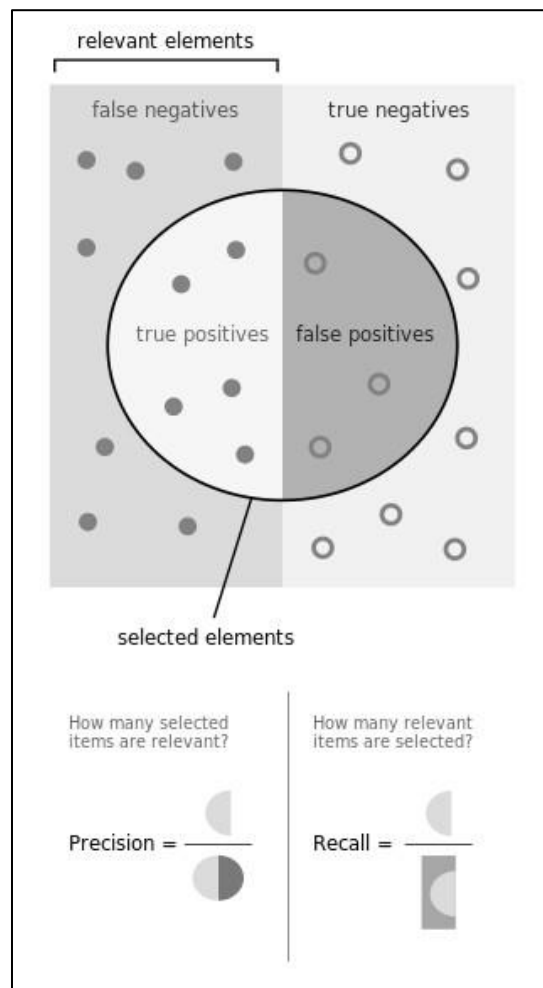
- Sensibilidad: proporción de empresas clasificadas como adquiridas o fusionadas que realmente lo son entre todas las empresas que cumplen estas características.

$$\textbf{Sensibilidad} = \frac{VP}{VP + FN} \quad (6)$$

- Precisión: proporción de empresas clasificadas como adquiridas o fusionadas que realmente lo son sobre el total que se detecta.

$$\textbf{Precisión} = \frac{VP}{VP + FP} \quad (7)$$

Los indicadores de sensibilidad y precisión resultan ya que nos proporcionan información sobre las observaciones relevantes y las observaciones seleccionadas [Ilustración 10]. De todos modos la precisión es el indicador más interesante para nuestro estudio ya que del total de empresas marcadas como ‘1’ nos indica cuales realmente son ‘1’. Desde un punto de vista práctico esta medida resulta interesante ya que con una precisión alta se disminuye el riesgo de la inversión.



*Ilustración 10 – Representación de sensibilidad (recall) y precisión.*

## 5. EXPERIMENTACIÓN

.



## Capítulo 6

### Evaluación de resultados.

El objetivo de este capítulo es comprobar la validez del modelo elaborado a través de las herramientas descritas en el apartado de evaluación del modelo. En este capítulo se presentan y analizan los resultados obtenidos, al haber realizado 100 experimentos con cada algoritmo los datos que se presentan son una media de todas las predicciones para los algoritmos. En el capítulo también se intentarán cubrir otros aspectos en la valoración del modelo como la comparación entre las técnicas utilizadas y la obtención de conclusiones de cara a la mejora del modelo.

#### 6.1 Resultados

En este apartado se presentan los resultados obtenidos por cada método en los escenarios definidos en el capítulo anterior. Para la evaluación de los algoritmos se seguirá el modelo de evaluación con las herramientas descritas.

LDA		Real	
		0	1
Predicción	0	98.075	433
	1	75.585	481

*Tabla 10 - Matriz de confusión para los datos de test aplicados al algoritmo de clasificación LDA. Los valores representados son la media de las clasificaciones para los 100 entrenamientos y predicciones.*

## 6. EVALUACIÓN DE RESULTADOS

KNN		Real	
		0	1
Predicción	0	93.926	377
	1	79.734	537

*Tabla 11 - Matriz de confusión para los datos de test aplicados al algoritmo de clasificación KNN. Los valores representados son la media de las clasificaciones para los 100 entrenamientos y predicciones*

SVMlin		Real	
		0	1
Predicción	0	145.980	744
	1	27.680	170

*Tabla 12 - Matriz de confusión para los datos de test aplicados al algoritmo de clasificación SVM linear. Los valores representados son la media de las clasificaciones para los 100 entrenamientos y predicciones*

SVMlog		Real	
		0	1
Predicción	0	105.377	436
	1	68.283	478

*Tabla 13 - Matriz de confusión para los datos de test aplicados al algoritmo de clasificación SVM gaussiano. Los valores representados son la media de las clasificaciones para los 100 entrenamientos y predicciones*

		Real	
		0	1
Predicción	0	98.216	353
	1	75.444	561

*Tabla 14 - Matriz de confusión para los datos de test aplicados al algoritmo de clasificación Random Forest. Los valores representados son la media de las clasificaciones para los 100 entrenamientos y predicciones*

Algoritmo	Moda Parametros	Acierto	Sensibilidad	Precisión
lda	-	0,5645 (0,0972)	0,5265(0,1064)	0,0063 (0,0003)
knn	k=61	0,5408 (0,0171)	0,5870 (0,0236)	0,0069 (0,0003)
svmLin	C=1000	0,8371 (0,167)	0,1864 (0,1701)	0,0065 (0,007)
svmLog	C=1000	0,6026 (0,0678)	0,5182 (0,0687)	0,0070 (0,0004)
rf	gamma=1	0,5657 (0,0079)	0,6138 (0,0196)	0,0074 (0,0001)
	ntree=500			
	mtry=2			
	nodesize=1			

*Tabla 15 – Resumen métodos de clasificación utilizados. Los valores acierto, sensibilidad y precisión se presentan con el formato media(desv.std.). Los valores representados son la media de las clasificaciones para los 100 entrenamientos y predicciones*

## 6.2 Evaluación de resultados

La tabla [Tabla 15] muestra un resumen comparativo de los distintos métodos estudiados, aunque los resultados son muy similares para todos los procedimientos: aciertos alrededor del 50%-60%, sensibilidad entre 50%-60% y precisión en torno al 0.7%. El único que no sigue este comportamiento es el svm lineal [Tabla 15]. Además, las desviaciones típicas pequeñas indican una cierta estabilidad entre las distintas simulaciones.

En vista de la tabla podemos afirmar que Random Forest parece ligeramente superior al resto de procedimientos. Pero, en líneas generales, podemos afirmar que todas las técnicas de clasificación **no llegan al rendimiento deseado** ya que la precisión (nuestro principal objetivo) es apenas superior al de realizar una clasificación aleatoria (0.5%).

Otra conclusión interesante es que los métodos sencillos utilizados como benchmark (LDA y KNN) obtienen resultados muy cercanos a los más complejos (SVM y RF). Esto nos

## 6. EVALUACIÓN DE RESULTADOS

lleva a la conclusión de que **el tipo de técnica empleada es poco relevante a la hora de la predicción con esta distribución de los datos.**

El objeto de predicción de este estudio (encontrar los '1') es la predicción de fusiones y adquisiciones de empresas. Por tanto, lo que nos interesa es maximizar la probabilidad de que las empresas marcadas como '1' sean realmente '1', minimizando al mismo tiempo los falsos positivos. Esta información es recogida por la *precisión* la cual es mayor para el algoritmo Random Forest con un total de 0,7%. Este valor es muy bajo teniendo en cuenta que una selección aleatoria al azar sería 0,5% y por tanto, no es adecuado para nuestro estudio.

El hecho de que todos los algoritmos se comporten de la misma forma, cuando el único factor que no varía en la predicción son los datos, nos lleva a la conclusión de que o bien el problema es extremadamente complejo o los datos utilizados no han sido los correctos. En este sentido pueden influir factores tales como el apuntado por Steven (1973): *“Cuando las variables son altamente colineales, los pesos en el modelo resultante son altamente inestables y el modelo pasa a ser totalmente dependiente de los datos introducidos como ejemplo por lo tanto la interpretación se convierte en una tarea muy difícil”*.

Hemos observado que la capacidad de predicción se mantiene bastante estable ante cambios en los parámetros y solo se ve afectada de forma significativa si cambiamos de composición las muestras<sup>2</sup>. Así, cambiando la proporción de 0's y 1's en el conjunto de

---

<sup>2</sup> Las predicciones de experimentos previos con diferentes distribuciones en los conjuntos de entrenamiento presentaron los siguientes resultados. Balanceo del conjunto entrenamiento 1:190 un '1' por 190 '0' para knn, svmLin, svmLog, rf (acierto  $\approx$  100%, sensibilidad  $\approx$  0%, precisión  $\approx$  0%) a partir de un conjunto de entrenamiento con este balanceo predice todos los datos como '0'. Balanceo del conjunto de entrenamiento 1:4 un '1' por 5 '0' para knn, svmLin, svmLog, rf (acierto  $\approx$  80%, sensibilidad  $\approx$  0%, precisión  $\approx$  0%) a partir de un conjunto de entrenamiento con este balanceo predice como mínimo el 80% de los datos como '0'.

entrenamiento obtenemos los siguientes niveles de acierto **1:1** ~50%, **1:4** ~80%, **1:190**~100%. Sin embargo, los valores de la precisión asociados a estas distribuciones de entrenamientos son 1:1~0,7%, **1:4** ~0,4%, **1:190**~0%. Esto demuestra que es necesario hacer algún tipo de procesamiento de los datos en este sentido para evitar que todo se clasifique en la clase mayoritaria.

Otro punto positivo es que aunque los resultados no son los esperados en términos de precisión, si se observa una ligera mejora sobre la clasificación aleatoria. Lo que nos permite albergar la esperanza de que se pueda seguir mejorando. En este sentido nuestros resultados mejorarían a otros en la literatura como Palepu (1986) que con datos de las mismas características con 163 empresas objetivo y 256 no objetivo llegó a un media de precisión de ~15%. Ya que nosotros en un contexto similar ('balanceado') nos aproximamos a niveles de precisión del 50%.

Finalmente, aunque hay indicios positivos queda mucho margen de mejora. En primer lugar, se puede plantear un análisis de calidad de los datos más exhaustivo y proponer otras variables para incluir al modelo. La inclusión de variables dummy ayudaría probablemente a mejorar la predicción, al añadir más información al modelo para una mejor clasificación. Sería conveniente realizar un análisis de variables para ver que variables pueden presentar colinealidad y realizar una reestructuración de las mismas. También habría que revisar la calidad de los datos y hacer una selección de variables más adecuada para la predicción, observar detenidamente todos los tipos de '0' que existen y tratar de adaptarlos para que las diferencias entre estos y los '1' sean más notables, pero sin perder sus características. Sería interesante realizar otras clasificaciones a partir de diferentes conjuntos de

---

## 6. EVALUACIÓN DE RESULTADOS

entrenamiento y de test, con una cantidad de datos más reducida y de mayor calidad, a partir de que esas predicciones empiecen a tener valor de clasificación se estudiaría la forma de extrapolarlas a un caso de mayor realismo. Una de las fuentes de complejidad del problema puede deberse a la existencia de distintos grupos o clusters por lo que podría ser interesante estudiar la existencia de los mismos para ajustar modelos diferentes en los distintos grupos.

## Capítulo 7

### Explotación del modelo

En esta sección del documento se explican las diferentes funcionalidades que puede tener el proyecto, dándole prioridad a la referida a inversión. Las dos primeras partes de la sección las conforman una introducción teórica al arbitraje y específicamente al arbitraje en operaciones de fusión y adquisición junto con unos escenarios ejemplo de cómo realizar las técnicas de arbitraje. La última parte indica otros usos que se le puede dar a modelo y como explotarlo.

#### 7.1 Arbitraje

El arbitraje en las operaciones de fusión y adquisición es una estrategia que de forma simultánea compra y vende acciones de dos empresas, generalmente involucradas en este tipo de operaciones.

El concepto de arbitraje se basa en la compra de valores de un mercado para posterior e inmediatamente revenderlos en otro mercado para obtener un beneficio derivado de la diferencia de precios. Desde el punto de vista de los fondos de inversión, el arbitraje más comúnmente se refiere a la absorción y venta simultánea de dos valores cuyos precios, bajo criterio del *trader*, no están en sintonía con lo que el trader cree que debe ser su *valor verdadero* (BarclayHedge, 2017). Actuando de acuerdo a la asunción de que los precios fluctuarán hacia su valor verdadero a lo largo del tiempo, el trader venderá en corto los valores con un precio sobreestimado y comprará los que lo tienen infra estimado. Una vez

## 7. EXPLOTACIÓN DEL MODELO

que los precios reviertan hacia su valor verdadero, la operación podrá ser liquidada y obtener beneficios de ella.

Para comprender la operación descrita con anterioridad es necesario entender el significado de vender en corto o estar en una *posición en corto*. Este no es más que simplemente adquirir un valor en concepto de préstamo, venderlo y posteriormente, esperando que su valor se vea decrementado antes de que el plazo del préstamo finalice, comprarlo por un valor más bajo y devolver el valor recibido en préstamo. Las *posiciones largas* son totalmente lo contrario, las utilizan los inversores cuando creen que el mercado va a subir y quieren sacar rentabilidad de ello. Un ejemplo sencillo es cuando un inversor cree que una determinada divisa se apreciará en el corto o largo plazo y por lo tanto compra una cantidad de la misma. Dicho inversor espera a que la divisa aumente su valor con el tiempo, cuando ocurre, podrá venderla a un precio mayor del que ha pagado por ella.

El arbitraje en fusiones y adquisiciones es un tipo de inversión guiada por eventos, cuyo objetivo es explotar la ineficiencia de precios que pueda ocurrir antes o después de una operación corporativa y que en el corto plazo produzca una mala valoración de los mismos.

Con el objetivo de ilustrar lo anteriormente dicho con un ejemplo, podemos considerar el caso de una operación de fusión o adquisición. Cuando una empresa muestra su intención de comprar a otra, el precio de la acción de la empresa objetivo normalmente sube y el precio de la adquisidora normalmente disminuye (BarclayHedge, 2017). Sin embargo, el precio de la acción de la empresa objetivo normalmente se mantiene por debajo del precio de adquisición, de este modo, el descuento presente en el precio indica la incertidumbre por parte del mercado sobre si la operación finalmente se llevará a cabo o no.



### 7.1.1 Estrategias de arbitraje en operaciones de fusión y adquisición

Como hemos identificado a lo largo del proyecto, en muchos casos, cuando se realizan operaciones de fusión o adquisición es posible como inversor obtener un beneficio derivado de ello. A continuación, se describirán las estrategias más importantes que buscan obtener beneficio del anuncio de estas operaciones y que aunque los mayores beneficios potenciales se producen cuando la operación aún no ha sido anunciada, es posible obtener ganancias incluso después del anuncio.

- Estrategia en largo: en una estrategia de arbitraje en fusiones y adquisiciones que solo involucra posiciones en largo, el inversor o trader compra acciones de la empresa objetivo de la operación si están valoradas con un descuento significativo a lo que se supone que la empresa adquisidora ofrecerá. Con las expectativas de que ese descuento se vaya haciendo más pequeño, hasta llegar a estar a la par del precio de adquisición, momento en el que al vender las acciones obtendrá un beneficio.
- Estrategia en largo y en corto: mientras que la estrategia en largo es típicamente usada por inversores minoristas, los inversores institucionales prefieren una estrategia de arbitraje que involucre ir en largo con las acciones de la empresa objetivo y en corto con las acciones de la empresa adquisidora. Las posiciones en corto con la empresa adquisidora se justifican porque a veces los adquisidores pagan de más por una adquisición atractiva. En ese caso, el precio de las acciones del adquisidor disminuirá significativamente, bien después del anuncio o a lo largo del tiempo. La combinación de posiciones en largo y posiciones en corto busca no solo obtener beneficio de la apreciación de las acciones de la empresa objetivo, sino también por una depreciación de las acciones de la empresa adquisidora.

## 7. EXPLOTACIÓN DEL MODELO

- Competencia u oferta más alta: en algunas situaciones, el precio de las acciones puede estar valorado por encima del precio de oferta en la adquisición. Esto puede ocurrir si los inversores piensan que el precio ofrecido por la acción de la empresa objetivo es demasiado bajo y que el adquirente bien se verá forzado a subir su oferta, o la empresa objetivo recibirá una oferta proveniente de otra empresa. Un inversor enfocado al arbitraje empezará por lo tanto con posiciones en largo en las acciones de la empresa objetivo o emitirá acciones de compra fuera de dinero suponiendo que el precio del ejercicio está por encima del precio del futuro subyacente.
- Fallo en la operación: en algunas ocasiones los traders pueden tener opiniones contrarias a la consecución de la operación, si piensan que la probabilidad de que sea completada es pequeña, bien por condiciones del mercado, por cuestiones regulatorias o porque el adquirente ha realizado una oferta hostil, etc. En este caso, un trader contrario a que se produzca la operación puede ponerse en corto con las acciones de la empresa objetivo, mientras compra acciones del adquirente.
- Arbitraje con empresas correlacionadas: esta estrategia de arbitraje a diferencia de las anteriormente descritas, que involucraban la empresa objetivo y la empresa adquirente, se basa en la realización de operaciones de compra y venta de acciones de empresas que están correlacionadas y que por lo tanto normalmente se encuentran en el mismo sector.

Esta es una estrategia de arbitraje parecida a la estrategia en largo pero que además busca proteger en el corto plazo las inversiones realizadas en las empresas que son supuestos objetivos de adquisición. Comparten la idea general de invertir en largo en las empresas que posteriormente serán adquiridas para vender las acciones

revalorizadas y de esta forma obtener beneficios. Al mismo tiempo que se invierte en esta empresa, se selecciona otra empresa del mercado que posea una alta correlación con ella y que por lo tanto presente un mismo comportamiento a cambios en el mercado. En esta nueva empresa seleccionada, que no tiene por qué ser la empresa adquirente se establecerán posiciones en corto para de esta forma cubrir las posiciones en largo establecidas en la empresa objetivo de adquisición. Al tener posiciones similares en corto y en largo de empresas altamente correlacionadas y que por lo tanto comparten un comportamiento similar en el mercado por acciones ajenas a ellas se pretende conseguir un saldo cero en el corto plazo. En el caso de que haya fluctuaciones en el mercado las dos empresas actuarán de forma similar y por lo tanto las posiciones en corto y en largo en ambas empresas se compensarán, esperando obtener resultado a través de una futura venta de las acciones en largo cuando se produzca la operación de adquisición.

## 7.2 Estrategias de inversión

Desarrollar una estrategia de inversión rentable es el objetivo final de este proyecto una vez se consigan unas predicciones con utilidad suficiente. Se pueden aplicar varias estrategias de inversión que exploten las ventajas del modelo de predicción, pero en este estudio nos centraremos en las estrategias de arbitraje *en largo* y con *empresas correlacionadas*.

Como ya se analizó en la *[¡Error! No se encuentra el origen de la referencia. ¡Error! No se encuentra el origen de la referencia.]* no sería posible llevar a cabo una estrategia de inversión con las predicciones obtenidas con este modelo. Por lo tanto, los ejemplos con los que trabajaremos a continuación se plantean bajo el supuesto de que el modelo presenta un 33,3% de precisión. Suponemos que se ha predicho que un total de tres empresas serán

## 7. EXPLOTACIÓN DEL MODELO

adquiridas este trimestre (01/01-31/03), para lo cual planteamos dos estrategias de inversión.

Como la formulación de un escenario real es prácticamente imposible, se ha decidido simular tres escenarios [Apéndice 4: Cotizaciones simuladas] con comportamientos similares correspondientes a diferentes sectores (aeronáutico, bancario, combustibles). En ellos se aplicarán las técnicas de arbitraje correspondientes y posteriormente se evaluarán los resultados obtenidos.

### 7.2.1 Arbitraje en largo

Las empresas A,B,C han sido clasificadas como adquiridas al final del cuatrimestre. Teniendo en cuenta que nuestro modelo tiene un 33% de acierto, al menos una de ellas será adquirida al final del cuatrimestre. Como estamos siguiendo una estrategia en largo, la estrategia será comprar acciones y en el momento que sus acciones aumenten procederemos a venderlas.

Los resultados han sido los siguientes:

Empresa	pAccion	01 - Enero		01 - Febrero		01-Marzo		31 - Marzo	
		Cot.	Rdto.	Cot	Rdto	Cot.	Rdto	Cot.	Rdto
A	12	12	0	12.9	0.9	13.21	1.21	13.75	1.75
B	20	20	0	19.28	-.72	19.96	-.04	18.66	-1.34
C	5	5	0	5.49	.049	5.68	0.68	9.1	4.1

*Tabla 16 – Estrategia en largo para tres empresas de diferentes sectores marcadas como ‘1’. La tabla muestra información de la evolución del precio de las acciones desde el momento de la compra en largo hasta el final del trimestre.*

*Rdto. Acumulado:*  $1.75 + (-1.34) + 4.1 = 4.51€$

**Bajo el supuesto de comprar 100 acciones de cada compañía.**

*Rdto. Acumulado:* 451€

*Total de la inversión:* 3.700€

*Rentabilidad:*  $(p_1 - p_0)/(p_0)$ : 0.122

### 7.2.2 Arbitraje con empresas correlacionadas

Las empresas A,B,C han sido clasificadas como adquiridas al final del cuatrimestre. Teniendo en cuenta que nuestro modelo tiene un 33% de acierto, al menos una de ellas será adquirida al final del cuatrimestre. Como estamos siguiendo una de arbitraje con empresas correlacionadas, la estrategia será comprar acciones de estas empresas y en el momento que sus acciones aumenten procederemos a venderlas. Para cubrirnos de posibles subidas y bajadas en el sector que estropeen los valores en largo que finalmente no son adquiridos venderemos en corto acciones de empresas altamente correlacionadas con las clasificadas para fusión ( $a^*, b^*, c^*$ ).

Los resultados han sido los siguientes:

Empresa	pAccion	01 - Enero		01 - Febrero		01-Marzo		31 – Marzo	
		Cot.	Rdto	Cot	Rdto	Cot.	Rdto	Cot.	Rdto
A	12	12	0	12.9	0.9	13.21	1.21	13.75	1.75
$a^*$	10.5	10.5	0	11.45	-.95	11.45	-.95	12	-1.5
B	20	20	0	19.28	-.72	19.96	-.04	18.66	-1.34
$b^*$	22	22	0	21.39	0.61	21.5	0.5	21	1
C	5	5	0	5.49	.049	5.68	0.68	9.1	4.1
$c^*$	6	6	0	6.67	-.67	6.75	-.75	6.5	-.5

*Tabla 17- Estrategia con empresas correlacionadas para 3 empresas de diferentes sectores. La tabla proporciona información sobre la evolución de los precios de las empresas con probabilidad de ser adquiridas y otras empresas altamente correlacionadas a ellas en el sector.*

## 7. EXPLOTACIÓN DEL MODELO

*Rdto. Acumulado:*  $(1.75 + (-1.5)) + ((-1.34) + 1) + (4.1 + (-.5)) = 3.51$

**Bajo el supuesto de comprar 100 acciones de cada compañía.**

*Rdto. Acumulado:* 702€

*Total de la inversión:* 7.550€

*Rentabilidad:*  $(p_1 - p_0) / (p_0)$ : 0.093

### 7.2.3 Evaluación estrategias

A priori la estrategia de inversión en largo sería la seleccionada ya que proporciona con una menor inversión, una mayor rentabilidad (~12%) mientras que la inversión con empresas correlacionadas (~9%).

No obstante, es necesario tener en cuenta que si se esperan tendencias bajistas en un sector la elección de estrategia óptima sería la correlacionada entre empresas ya que al invertir en ambas empresas en posiciones contrarias una operación amortigua a la otra.

## Capítulo 8

### Planificación y presupuesto

En este apartado se detallan los datos relacionados con la planificación del tiempo y costes del proyecto para poder llevarlo a cabo. En el apartado de planificación se proporcionarán detalles relacionados con la estimación de tiempo para cada fase de proyecto y el tiempo finalmente incurrido a través de un diagrama de Gantt. En la sección de presupuesto, se determinará el coste del proyecto para el tiempo inicialmente planificado.

#### 8.1 Planificación

La estimación del tiempo para este proyecto se ha realizado imputando a cada etapa de la metodología aplicada la cantidad de horas estimadas. Posteriormente, se comparan los resultados estimados con los reales, observando si finalmente se han cumplido los plazos o si se ha incurrido en algún retraso.

Ha sido necesario para completar la asignación del tiempo a cada tarea complementar la metodología con algunas fases que no aparecen explícitamente indicadas en el desarrollo del proyecto.

##### *8.1.1 Asignación de horas estimadas*

La asignación se realizará en días de trabajo. 1 día de trabajo equivale a 8h. El proyecto comienza el día 20/01/2017 y termina el 14/06/2017 el equivalente a 145 días de trabajo.

## 8. PLANIFICACION Y PRESUPUESTO

- **Comprensión del negocio:** en esta fase se presentan los objetivos y requisitos del proyecto. Supone el inicio del proyecto, entender el problema a resolver, la comprensión de objetivos y elección de las herramientas que darán soporte. Comprende también una revisión de la literatura. Tiempo estimado asignado: 15 días.
- **Comprensión de los datos:** esta fase es el primer contacto con los datos, en ella se realiza una comprensión de los mismos, entendiéndolos de una manera global y se realizan análisis para tener una idea fundamentada de la distribución de los mismos. Tiempo estimado asignado: 20 días.
- **Preparación de los datos:** se toman decisiones en relación a la estructura y formato de los datos. En esta fase se deciden hacer cambios de formato, eliminación de variables y normalización de los datos. Tiempo estimado asignado: 20 días.
- **Modelado:** esta fase ha sido la más larga de todo el proyecto ya que se ha intentado ajustar el modelo para sacar el máximo rendimiento de los datos disponibles. Tiempo estimado asignado: 40 días.
- **Evaluación:** esta fase ha ido de la mano con el modelado, comprende la comparación entre algoritmos de clasificación y evaluación de la calidad de los resultados. Tiempo estimado asignado: 25 días.
- **Despliegue:** elaboración de un benchmark para ilustrar la explotación económica del proyecto. Tiempo estimado asignado: 5 días.



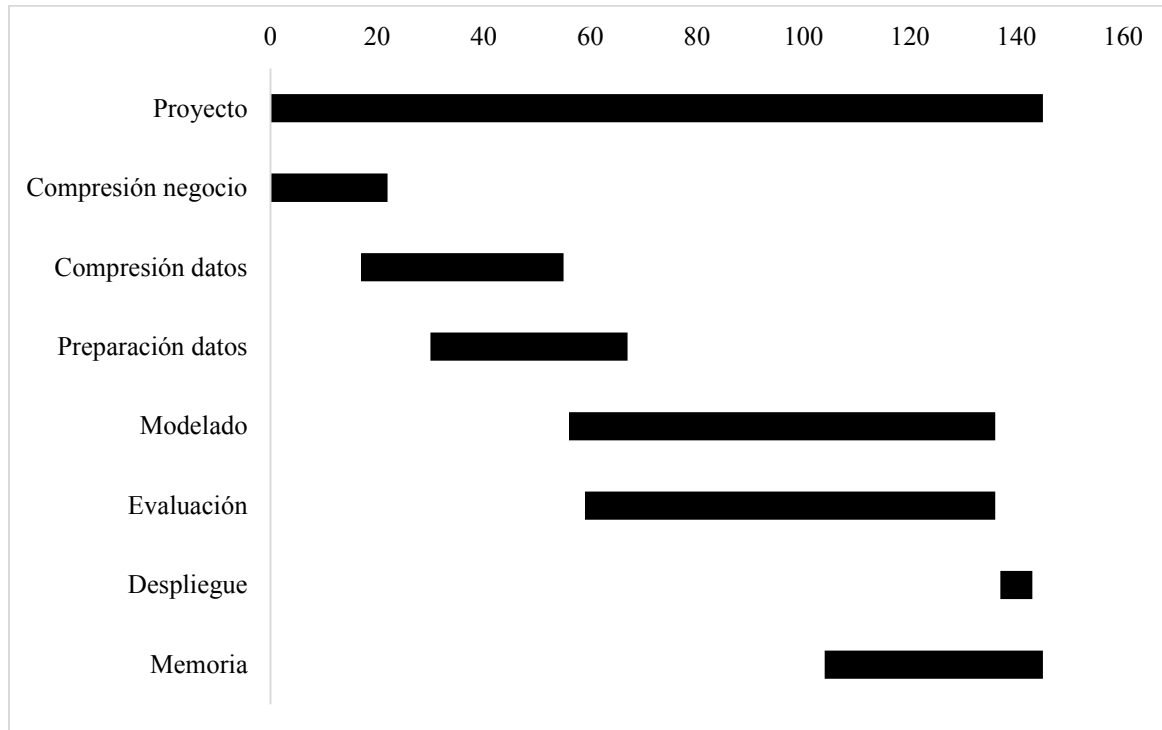
- Memoria: el documento de la memoria se ha ido redactando paralelamente al resto del trabajo conforme los avances se fueron consiguiendo. Tiempo estimado asignado 19 días.

	Fecha Inicio	Fecha Fin	Duración*	Espera	Días estimados
<b>Proyecto</b>	20/01/2017	14/06/2017	145	0	145
<b>C. Negocio</b>	20/01/2017	11/02/2017	22	0	15
<b>C. Datos</b>	06/02/2017	16/03/2017	38	17	20
<b>Prep. Datos</b>	19/02/2017	28/03/2017	37	30	20
<b>Modelado</b>	17/03/2017	05/06/2017	80	56	40
<b>Evaluación</b>	20/03/2017	05/06/2017	77	59	25
<b>Despliegue</b>	06/06/2017	12/06/2017	6	137	5
<b>Memoria</b>	04/05/2017	14/06/2017	41	104	19

*Tabla 18 – Detalles Horas estimadas. Se proporcionan la fecha de inicio y fin de cada etapa del proyecto. La duración está expresada en días reales, se cuenta como un día si durante ese día se le dedicó tiempo al proyecto, independientemente de las horas empleadas.*

El gráfico Gantt quedaría de la siguiente manera, con un total de 145 días que equivalen a 1.160 h estimadas.

## 8. PLANIFICACION Y PRESUPUESTO



*Tabla 19 – Diagrama Gantt Planificación. Se muestra el tiempo de dedicación previsto para las etapas del proyecto. . El proyecto comienza el día 20/01/2017 y termina el 14/06/2017 el equivalente a 145 días de trabajo*

### 8.2 Presupuesto

Una vez terminada la planificación temporal del proyecto, es necesario estimar el presupuesto del mismo. El presupuesto es una parte fundamental a tener en cuenta a la hora de decidir si llevar a cabo un proyecto. En este caso y aun tratándose de un proyecto pensado para invertir en bolsa hemos decidido separar los conceptos de desarrollo y despliegue a la hora de calcular el impacto económico. De este modo tendremos un mayor detalle de cuanto ha resultado el coste en investigación y cuál es el rendimiento del modelo.

A continuación, presentamos el presupuesto para las 1.160 h de trabajo estimado, hemos decidido dividirlo en tres grupos: coste del personal, coste del equipamiento y coste del software.

<b>Personal</b>				
<b>Nombre</b>	<b>Puesto</b>	<b>Tarifa €/h</b>	<b>Horas</b>	<b>Coste total</b>
Sebastián Freire García	Investigador	35	1.160	40.600€
<b>TOTAL</b>				<b>40.600€</b>

*Tabla 20 – Presupuesto coste del personal.*

<b>Equipamiento</b>				
<b>Descripción</b>	<b>Duración</b>	<b>Amortización</b>	<b>Coste</b>	<b>Coste Imputable</b>
HP Pavillion dv6	5,06 meses	48 meses	999€	105,31€
<b>TOTAL</b>				<b>105,31€</b>

*Tabla 21 – Presupuesto coste del equipamiento.*

<b>Software</b>				
<b>Descripción</b>	<b>Duración</b>	<b>Amortización</b>	<b>Coste</b>	<b>Coste Imputable</b>
Lincencia Microsoft Office 2013	5,06 meses	36 meses	1.200€	168,66€
<b>TOTAL</b>				<b>168,66€</b>

*Tabla 22 – Presupuesto coste del software.*

En las tablas anteriores están incluidos aquellos recursos necesarios para la etapa del desarrollo del proyecto que implican algún coste. No obstante, a lo largo del desarrollo se han utilizado diversos recursos que no se han incluido en el presupuesto al no implicar

## 8. PLANIFICACION Y PRESUPUESTO

coste. Es importante distinguir todos los recursos utilizados en el desarrollo del proyecto, tanto físicos como herramientas software e identificarlos en la lista de recursos utilizados. Finalmente, se añade un resumen de los costes incurridos durante la realización del proyecto.

<b>Resumen</b>	
Concepto	Coste
Personal	40.600€
Equipamiento	105,31€
Software	168,66€
Total sin IVA	40.837,97€
IVA (21%)	8.538,53€
<b>TOTAL</b>	<b>49.421,5€</b>

*Tabla 23 – Presupuesto del coste total del proyecto.*

## Capítulo 9

### Marco regulador

Para este proyecto el marco regulador se puede ver desde dos enfoques diferentes: en la operativa o hacia el despliegue. El primero comprende la regulación, formas de adquisición, impuestos y problemas contables en los que puede ocurrir una empresa cuando realiza operaciones de absorción. Es importante tener estos conceptos en cuenta ya que pueden cambiar totalmente una predicción realizada. El segundo está relacionado con la legislación aplicada al modelo descrito en el estudio, trata sobre los requisitos necesarios para el diseño y la explotación.

#### 9.1 Operativa

##### 9.1.1 Legislación

Las operaciones de fusión y adquisición en Estados Unidos pueden ser rechazadas por las leyes federales antimonopolio. La más importante es el *Clayton Act 1914*, la cual prohíbe la adquisición cuando (States, 1914): “*en alguna línea de negocio o en alguna sección del país pueda ser sustancialmente para disminuir la competencia o para intentar crear un monopolio.*”

La ley antimonopolio puede ser aplicada por el gobierno federal de dos modos: por una demanda civil presentada por el Departamento de Justicia o por un procedimiento iniciado por la Comisión Federal de Comercio (FCT) (Brealey et al., 2013). El *Hart-Scott-Rodino Antitrust Act* obliga a informar de operaciones de acciones de que involucre \$15 millones

## 9. MARCO REGULADOR

o el 15% de la empresa objetivo. Así, casi todas las grandes fusiones son revisadas en una etapa temprana (Brealey et al., 2013). Tanto el departamento de justicia como la FCT poseen medidas cautelares para retrasar una fusión.

### 9.1.2 *Formas en la adquisición*

Suponiendo que se cumplen todos los requisitos de la ley antimonopolio una posibilidad es *fusionar* ambas empresas, por lo que una compañía asume *todos* los *activos* y todas las *deudas* de la otra. Una fusión de este tipo ha de tener la aprobación de al menos el 50% de los accionistas de cada compañía (Robert F. Bruner, 2011).

Una alternativa, es comprar todas las acciones de la empresa en efectivo, acciones u otros valores. En esto caso el comprador puede negociar individualmente con los accionistas de la compañía adquirida y la dirección no tiene por qué estar involucrada en absoluto. Su aprobación y cooperación es deseada, pero en el caso de que se opongan, el comprador intentará hacerse con la mayoría de las acciones de los accionistas y una vez obtenido el control puede completar la fusión.

El tercer enfoque es comprar activos de la empresa deseada. En ese caso la posesión de los activos necesita ser transferida y el pago es realizado a la compañía en vez de directamente a los accionistas.

### 9.1.3 *Contabilidad*

Desde 2001, debido a los ajustes realizados por el *Financial Accounting Standards Board* (FASB) se requiere que la compañía adquisidora use el método de contabilidad de la empresa adquirida (Ramanna & Allen, 2016).

#### 9.1.4 Impuestos

Una adquisición puede ser tasada o estar libre de impuestos. Si el pago se realiza en efectivo, la adquisición es considerada como tasable. En este caso a los accionistas de la empresa comprada se les trata como que han tenido que *vender* sus acciones, y por lo tanto deben pagar impuestos por los rendimientos que les han proporcionado (Brealey et al., 2013). Si el pago es efectuado en forma de acciones, la adquisición está libre de impuestos y se considera que los accionistas han *intercambiado* sus antiguas acciones por unas nuevas similares, no se reconocen ni beneficios ni pérdidas (Brealey et al., 2013).

### 9.2 Despliegue

En el momento del despliegue o de la explotación del modelo, en caso de quererlo llevar a la práctica, sería importante tener en cuenta algunas consideraciones como la legislación vigente a la hora de invertir y la disponibilidad de los datos para conseguir un sistema más robusto.

Al tratarse de empresas que operan en el mercado estadounidense se han de seguir las reglas de inversión correspondientes al sector americano. Estas son aplicadas por la SEC, *U.S Security and Exchange Comission*, que es una agencia independiente del gobierno federal de los Estados Unidos. El SEC se encarga principalmente de reforzar las leyes federales de seguridad, proponer nuevas leyes y regular el mercado de valores, los títulos de valor fijo estatal y de supervisar la seguridad en los mercados de alta frecuencia.

Por otro lado, los datos se han obtenido de la base de datos de Compustat (Standard & Poor's) y de SDC Platinum (Thompson Reuters). Estas bases de datos no son de acceso público por lo tanto si se desea adquirir más datos o actualizar el panel de datos actual sería

## 9. MARCO REGULADOR

necesario cerciorarse de que la licencia de uso de la base de datos está activa. No obstante, con un correcto modelado los datos se podrán extraer de otras fuentes.



## Capítulo 10

### Conclusiones y trabajo futuro

#### 10.1 Conclusiones

El problema de clasificación planteado resulta un problema interesante ya que hasta el momento no se había abordado la predicción de operaciones de fusión y adquisición desde el punto de vista del aprendizaje automático. Las conclusiones obtenidas de este trabajo son de dos tipos: relacionadas con el aprendizaje automático y financieras.

Debido a la complejidad computacional que conlleva es un problema muy difícil de resolver. Además, plantea muchas dificultades ya que no se conocen con claridad los datos que mejor lo predicen.

El problema presentaba un extremado desequilibrio en los datos, lo cual, para técnicas de aprendizaje automático representa un problema a la hora de rendir adecuadamente.

Con la utilización de variables financieras para la predicción (*atq*, *ceqq*, *dltqq*, *niq*, *saleq*, *per*, *roe*, *leverage*) los métodos estándar no han dado el rendimiento esperado, aunque si se pueden observar algunos indicios positivos. El error en la clasificación es menor que el de una clasificación al azar, por lo tanto, bien mejorando los datos o bien mejorando la metodología se podrían mejorar los resultados.

La información obtenida de la predicción no sería eficiente en una extrapolación al mundo real aplicada a la gestión de una cartera de acciones.

## **10.2 Trabajo futuro**

A continuación, se describirán algunas líneas de trabajo futuro con las que mejorar el modelo planteado en el estudio.

En primer lugar, a partir de los datos actuales se podrían crear otros indicadores. A través de un análisis de atributos podríamos eliminar las variables que no aportan información al modelo y crear otras basadas en ratios financieros que la literatura respalde como significativos. Una configuración del panel con variables donde se estudie la evolución de la compañía en el tiempo podría aportar mucha información para la clasificación.

Segundo, con referencia al tipo de entrenamiento se podrían plantear otros algoritmos de clasificación como las redes neuronales y una optimización más exhaustiva de los ya utilizados.

En tercer lugar, añadir al panel de datos variables de mercado, de tipo tecnológico o macroeconómicas aportaría una visión más global del problema, ya que no son únicamente los motivos económicos los que generan este tipo de operaciones.

En cuarto lugar, la búsqueda de clusters o grupos de empresas con características similares para poder realizar la clasificación dentro de estos subgrupos.

Por último, aparte de únicamente estudiar las potenciales empresas objetivo, hacer hincapié también en las empresas que potencialmente puedan realizar una oferta. De este modo si

conseguimos ver similitudes entre empresas de las cuales una es potencial objetivo y otra potencial compradora tendríamos más información sobre una posible absorción.



## Lista de referencias

- Ambrose, B. W. ., & Megginson, W. (2011). The Role of Asset Structure , Ownership Structure , and Takeover Defenses in Determining Acquisition. *Business*, 27(4), 575–589.
- Andrade, G., Mitchell, M., & Stafford, E. (2001). New Evidence and Perspectives on Mergers. *Journal of Economic Perspectives*, 15(2), 103–120.  
<https://doi.org/10.1257/jep.15.2.103>
- Baillo, A., & Cuevas, A. (2008). Supervised functional classification: A theoretical remark and some comparisons, 18. Retrieved from <http://arxiv.org/abs/0806.2831>
- BarclayHedge. (2017). Understanding Merger Arbitrage. Retrieved from <https://www.barclayhedge.com/research/educational-articles/hedge-fund-strategy-definition/hedge-fund-strategy-merger-arbitrage.html>
- Barnes, P. (2000). The identification of U.K. takeover targets using published historical cost accounting data: Some empirical evidence comparing logit with linear discriminant analysis and raw financial ratios with industry-relative ratios. *International Review of Financial Analysis*, 147–162.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. (M. Jordan, J. Kleinberg, & B. Scho, Eds.).
- Bradley, M., & Kim, E. H. (1988). Synergistic gains from corporate acquisitions and their division between the stockholders of target and acquiring firms, 21, 3–40.
- Brealey, R. A., Myers, C. S., & Allen, F. (2013). *Principles of Corporate Finance*.
- Breiman, L. (2001). Random Forests. *Statistics Department University of California*, 1–33.
- Bruner, R. F. (2004). *Applied Mergers and Acquisitions*.
- Bruner, R. F. (2011). Applied Mergers and Acquisitions, (January), 384. Retrieved from [https://books.google.com/books?id=BigsPJZ0E\\_4C&pgis=1](https://books.google.com/books?id=BigsPJZ0E_4C&pgis=1)

## LISTA DE REFERENCIAS

- Cherkassky, V., & Mulier, F. M. (2007). *Learning from data. Concepts, theory and methods*. <https://doi.org/10.1198/tech.2001.s558>
- Cremers, K. J. M., Nair, V. B., & John, K. (2009). Takeovers and the Cross-Section of Returns. *The Review of Financial Studies*, 22(4), 1409–1445. <https://doi.org/10.1093/rfs/hhn032>
- Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press.
- Devroye, L., Györfi, L., & Lugosi, G. (2014). A Probabilistic Theory of Pattern Recognition. *Igarss 2014*, 31(1), 1–5. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Di Deco Sampedro, J., & Díaz García, J. (2012). Estudio y aplicación de técnicas de aprendizaje automático orientadas al ámbito médico: estimación y explicación de predicciones individuales, 103. Retrieved from [https://repositorio.uam.es/bitstream/handle/10486/12100/59264\\_Di\\_Deco\\_Sampedro\\_JavierPFM.pdf?sequence=1](https://repositorio.uam.es/bitstream/handle/10486/12100/59264_Di_Deco_Sampedro_JavierPFM.pdf?sequence=1)
- Dietrich, J. K., & Sorenson, E. (1984). An Application of Logit Analysis to Prediction of Merger Targets. *Journal of Business Research*, 393–402.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Pattern Classification. *New York: John Wiley, Section*. <https://doi.org/10.1007/BF01237942>
- Duke Energy Corporation. (2012). Duke Energy, Progress Energy Complete Merger. Retrieved from <https://news.duke-energy.com/releases/duke-energy-progress-energy-complete-merger>
- Eckbo, B. E. (2009). Bidding strategies and takeover premiums: A review. *Journal of Corporate Finance*, 15(1), 149–178. <https://doi.org/10.1016/j.jcorpfin.2008.09.016>
- Fama, E. F., & French, K. R. (1992). The Cross-section of Expected Stock Returns.
- Fix, E., & Hodges, J. L. (1951). An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation.
- Froese, H. G. (2013). Predicting Takeover Targets.
- Harris, R. S., Stewart, J. F., Guilkey, D. K., & Willard T., C. (1982). Characteristics of

- Acquired Firms: Fixed and Random Coefficients Probit Analyses. *Southern Economic Journal*, 49(1), 164–184.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*.
- Horning, N., Liaw, A., Wiener, M., Qian, Y., Zhou, W., Yan, J., ... Zhang, X. (2014). Support Vector Machine Classification of Object-Based Data for Crop Mapping , Using Multi-Temporal Landsat Imagery. *R News*, XXXIX(1), 54.1-54.10. <https://doi.org/10.5244/C.22.54>
- Ihaka, R. (1998). R: Past and Future History. *Proceedings of the 30th Symposium on the Interface*, 392–396. Retrieved from <http://cran.r-project.org/doc/html/interface98-paper/paper.html>
- Jarrell, G. A., & Poulsen, A. B. (1989). The Returns to Acquiring Firms in Tender Offers: Evidence from Three Decades.
- Langestieg, T. . (1978). An application of a three-factor performance index to measure stockholder gains from mergers.
- Marbán, O., Mariscal, G., & Segovia, J. (2009). *Data Mining and Knowledge Discovery in Real Life Applications*. (J. Ponce & A. Karahoca, Eds.). I-Tech Education and Publishing.
- Max Kuhn Contributions from Jed Wing, A., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., ... Max Kuhn, M. (2017). Package “caret.” Retrieved from <https://cran.r-project.org/web/packages/caret/caret.pdf>
- Meador, A. L., Pamela, C. H., & Gayle, R. L. (1996). Development of Prediction Models for Horizontal and Vertical Mergers. *Journal of Financial and Strategic Decision*, 9(1), 11–23.
- Melicher, R. W., & Nielsen, J. F. (1977). Financial Factors That Affect Acquisition Prices Melicher, Ronald W;Nielsen, James F Review of Business and Economic Research; Winter 1977; 13, 2; ProQuest pg. 95.

## LISTA DE REFERENCIAS

- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., & Lin, C.-C. (2017). *R Package e1071 Version 1.6-8. Gpl-2*. Retrieved from <https://cran.r-project.org/web/packages/e1071/e1071.pdf>
- Mitchell, M. L., & Mulherin, J. H. (1996). The impact of industry shocks on takeover and restructuring activity. *Journal of Financial Economics*, 193–229.
- Palepu, K. G. (1986). Predicting takeover targets. A methodological and empirical analysis. *Journal of Accounting and Economics*, 8(1), 3–35. [https://doi.org/10.1016/0165-4101\(86\)90008-X](https://doi.org/10.1016/0165-4101(86)90008-X)
- Pérez Porto, J., & Gardey, A. (2009). Definicion.de: Definición de Excel.
- Ragothaman, S., Naik, B., & Ramakrishnan, K. (2003). Predicting corporate acquisitions: An application of uncertain reasoning rule induction. *Information Systems Frontiers*, 401–412.
- Ramanna, K., & Allen, D. (2016). The Quiet War on Corporate Accountability. Retrieved from [https://www.nytimes.com/2016/04/26/opinion/the-quiet-war-on-corporateaccountability.html?smid=pl-share&\\_r=0](https://www.nytimes.com/2016/04/26/opinion/the-quiet-war-on-corporateaccountability.html?smid=pl-share&_r=0)
- Ripley, B., & Venables, W. (2015). Package “class”: Various Functions for Classification, Including k-Nearest Neighbour, Learning Vector Quantization and Self-Organizing Maps. *Cran*, 19. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4/>
- Ripley, W. N. V. and B. D. (2017). Package “MASS.”
- Rossi, S., & Volpin, P. F. (2004). Cross-country determinants of mergers and acquisitions. *Journal of Financial Economics*, 74(2), 277–304. <https://doi.org/10.1016/j.jfineco.2003.10.001>
- Rouse, M. (2017). Techtarget. Retrieved from <http://searchdatacenter.techtarget.com/es/definicion/Aprendizaje-automatico-machine-learning>
- Schwert, G. W., Barclay, M., Bradley, M., Brickley, J., Deangelo, L., French, K., ... Zimmerman, J. (1996). Markup Pricing in Mergers and Acquisitions.
- Simkowitz, M. A., & Monroe, R. . (1971). A Discriminant Analysis Function for



- Conglomerate Targets. *Southern Journal of Business*, 1–14.
- States, U. (1914). That the proceeds arising from the sale of the property described be covered into the Treasury of the United States as a miscellaneous receipt . Approved, October 14, 1914 .
- Stevens, D. L. (1973). Financial Characteristics of Merged Firms. *Journal of Financial and Quantitative Analysis*, 149–158.
- Tobergte, D. R., & Curtis, S. (2013). *Handbook of Corporate Finance - Empirical Corporate Finance Vol 1 & 2*.
- Tsagkanos, A., Georgopoulos, A., & Siriopoulos, C. (2007). Predicting Greek mergers and acquisitions: A new approach. *International Journal of Financial Services Management*, 289–303.
- Walkling, R., & Edminster, R. O. (1985). Determinants of Tender Offer Premiums. *Financial Analytics Journal*, 27–37.
- Wansley, J. . (1984). Discriminant Analysis and Merger Theory. *Review of Business and Economic Research*.
- Yang, C. (2014). Exploiting Technological Indicators for Effective Technology Merger and Acquisition ( M & A ) Predictions. *A Journal of the Decision Science Institute*, 22(1), 187–200.

## LISTA DE REFERENCIAS

## Apéndice

### Apéndice 1: Detalle del panel de datos

ID	BBDD	Longitud	Type
Gvkey	Compustat	6	character
Lpermno	Compustat	5	numeric
Datadate	Compustat	10	Date
Fyearq	Compustat	4	numeric
Fqtr	Compustat	3	numeric
Indfmt	Compustat	4	character
Consol	Compustat	1	character
Popsrc	Compustat	1	character
Datafmt	Compustat	3	character
Cusip	Compustat	9	character
Curedq	Compustat	3	character
Datacqtr	Compustat	6	character
Datafqtr	Compustat	6	character
Atq	Compustat	11	numeric
Ceqq	Compustat	10	numeric
Dlttq	Compustat	10	numeric
Niq	Compustat	10	numeric
Saleq	Compustat	10	numeric
Uniamiq	Compustat	9	numeric
Epspiy	Compustat	8	numeric
Costat	Compustat	1	character
Prccq	Compustat	16	numeric
Sic	Compustat	4	numeric
cusip6	Compustat	6	character
Qdate	Compustat	3	numeric
Size	Compustat	19	numeric
Roe	Compustat	21	numeric
Firm	Compustat	5	numeric
Growth	Compustat	20	numeric
Per	Compustat	21	numeric
Leverage	Compustat	20	numeric
_merge	Compustat	6	labelled
DateAnnounced	SDC Platinum	10	Date
DateEffective	SDC Platinum	10	Date
TargetName	SDC Platinum	30	character

## APENDICE

AcquirorName	SDC Platinum	30	character
ofSharesAcq	SDC Platinum	5	numeric
AcquirorCUSIP	SDC Platinum	6	character
AcquirorPrimarySICCode	SDC Platinum	4	character
ValueofTransactionmil	SDC Platinum	10	numeric
TargetCUSIP	SDC Platinum	6	character
MasterDealType	SDC Platinum	2	character
Form	SDC Platinum	16	character
TargetUltimateParentCUSIP	SDC Platinum	6	character
AcquirorUltimateParentCUSIP	SDC Platinum	6	character
ofCash	SDC Platinum	5	numeric
ofStock	SDC Platinum	5	numeric
Year	SDC Platinum	4	numeric
Month	SDC Platinum	2	numeric
Q	SDC Platinum	1	numeric
Merger		1	numeric

*Tabla 24 – Características del panel de datos. Esta tabla nos informa de todos los atributos que conforman cada observación. Representa la unión de las observaciones recopiladas de Compustat que nos informan de las características financieras de la empresa y de SDC Platinum que nos informan sobre la operación de fusión o adquisición.*

**Gvkey:** Global Company Key. Número único de seis dígitos asignado a una compañía (incidencia, divisa, índice) en la base de datos Capital IQ de Compustat. Representa la clave primaria para una compañía que forma parte del índice. La gvkey puede ser usada para seguir a una compañía a lo largo del tiempo, mientras que el nombre de la compañía, CUSIP, o ticker pueden cambiar.

**Impermno:** Es un número de identificación único y permanente asignado por el CRSP a cada seguridad. A diferencia del CUSIP, Ticker Symbol, Nombre de la Compañía, el PERMNO nunca cambia durante la historia, ni tampoco es reasignado.

**Datadate:** yyyy-mm-dd en el momento de la toma de datos.

**Fyearq:** Año fiscal. Periodo de 12 meses usado para calcular informes financieros anuales en negocios y otras organizaciones.

**Fqtr:** Fiscal quarter. Un periodo de tres meses en el calendario financiero que establece las bases del reporte de beneficios y el pago de dividendos.

**Indfmt:** Informa el tipo de industria. Puede ser FS (Servicios Financieros) u INDL (Industrials).

**Consol:** Información sobre la empresa padre y sus subsidiarias tratadas como una sola.

**C:** Consolidada. Empresas internacionales y de America del Norte.

**I:** Incidencias a nivel de fundamentales. Empresas internacionales

**N:** No consolidadas. Empresas internacionales.

**P:** Pre-FASB. Empresas de América del Norte.

**D:** Pre-Desinversión. Empresas de América del Norte.

**E:** Post- Desinversión. Empresas de América del Norte.

**R:** Pro-Forma. Empresas de América del Norte.

**Popsrc:** fichero de la fuente de población.

**datafmt:** formato fecha. STD (Estándar).

**cusip:** Identificador alfanumérico de nueve caracteres asignado a un activo financiero por una agencia independiente. Los seis primeros caracteres del CUSIP son identificadores de la compañía, los dos siguientes dígitos identifican un bien particular y el último dígito es empleado para mejorar la precisión de las transmisiones electrónicas de los CUSIP.

En COMPUSTAT, los seis primeros caracteres son equivalentes al CNUM (CUSIP issuer code) y los últimos tres son el CIC (CUSIP issue code). Durante la mayoría de la historia de COMPUSTAT, el CNUM ha sido usado como un identificador de la compañía, pero ha sido recientemente sustituido por el gvkey.

**curedq:** Estándar Internacional para los códigos de divisas.

**datacqtr:** año y quarter correspondiente a la observación.

**datafqtr:** año fiscal y quarter correspondiente a la observación.

**atq:** Cantidad total de activos que posee la empresa.

**ceqq:** Capital Social de la empresa.

**dlttq:** Deuda a largo plazo que consiste en préstamos y obligaciones financieras a más de un año.

**niq:** Beneficio después de intereses e impuestos.

**saleq:** Ingresos totales de la compañía, ingresos en efectivo y otros ingresos.

## APENDICE

**uniamiq:** Beneficio neto antes de gastos extraordinarios y después de interés no controlado.

**epspiy:** Ingresos por acción.

**costat:** Estado de la empresa en el mercado. Activo/Inactivo

**prccq:** Precio de la acción al finalizar el trimestre.

**sic:** Standard industry classification Code.

**cusip6:** Seis primeros dígitos del CUSIP.

**qdate:**

**size:** tamaño de la empresa.

**roe:** ratio de rentabilidad financiera. Relaciona el beneficio económico con los recursos necesarios para obtener ese lucro.

**Firm:** grupo del cusip al que pertenece.

**growth:** ratio de crecimiento de la empresa.

**per:** Ratio precio-beneficio.

**Leverage:** nivel de apalancamiento de la empresa

**Merge\_:** No aporta información.

**DateAnnounced:** Fecha anuncio de la operación.

**DateEffective:** Fecha efectiva de la operación.

**TargetName:** Nombre de la compañía que es adquirida.

**AcquirorName:** Nombre de la compañía adquisidora.

**ofSharesAcq:** Porcentaje total de la empresa que se ha adquirido.

**AcquirorCUSIP:** CUSIP de la empresa adquisidora.

**AcquirorPrimarySICCode**

**ValueofTransactionmil:** Valor de la operación.

**TargetCUSIP:** CUSIP de la empresa adquirida.

**MasterDealType:** Código de la operación.

**Form:** Tipo de operación.

**TargetUltimateParentCUSIP:** Cusip padre del target.

**AcquirorUltimateParentCUSIP:** Cusip padre del adquirente.

**ofCash:** Porcentaje total de la empresa adquirida en cash.

**ofStock:** Porcentaje total de la empresa adquirida en acciones.

**year:** Año en el que se produce la operación

**month:** Mes en el que se produce la operación

**q:** Trimestre en el que se produce la operación.

**Merger:** informa si la empresa ha sido adquirida '1' o si no ha sido adquirida '0'.

## Apéndice 2: Flujo del proyecto

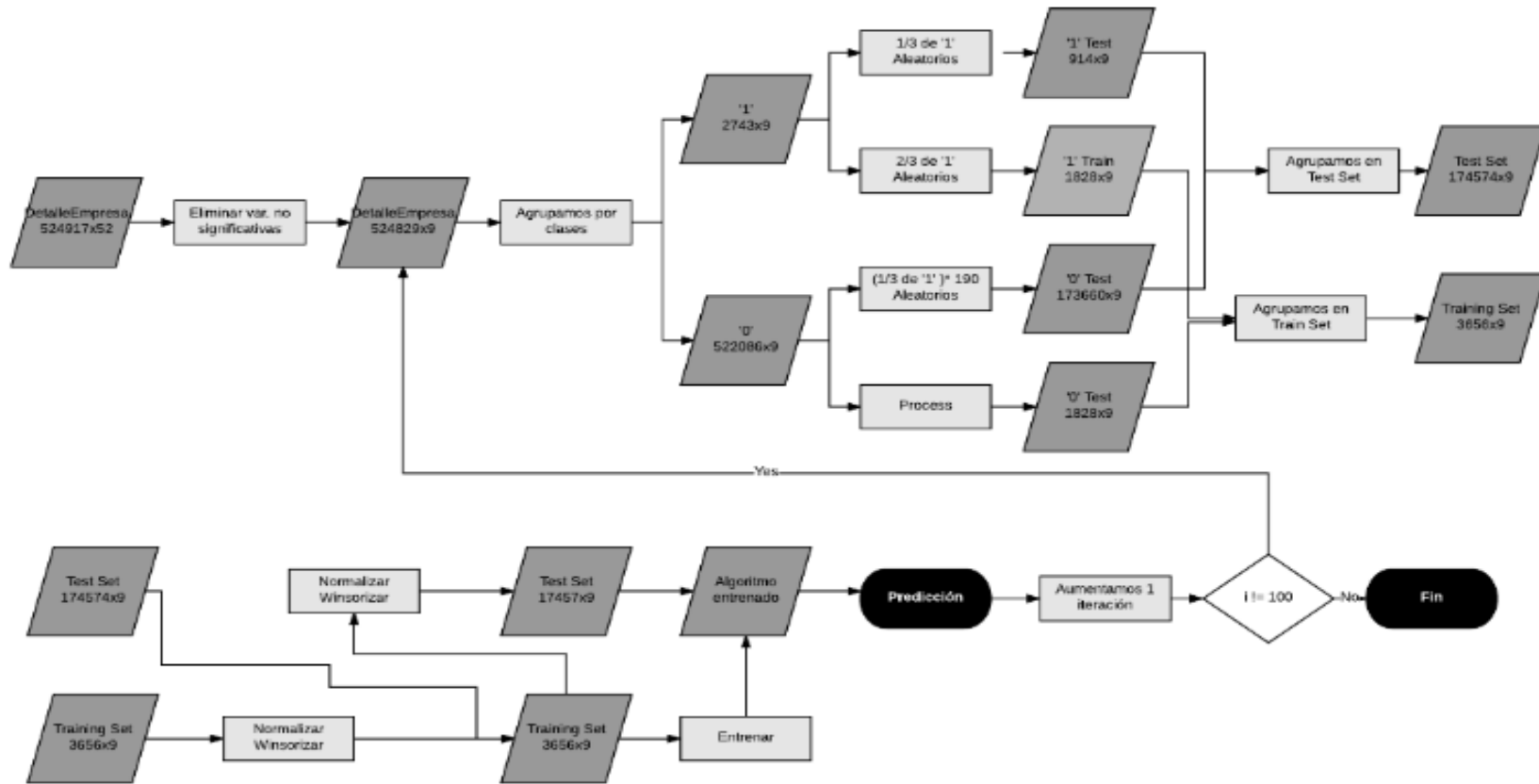


Ilustración 11 – Flujo de programación del proyecto. Es un esquema conceptual sobre las operaciones que se realizan en el proyecto desde el procesamiento de datos hasta la predicción de resultados.



### Apéndice 3: Aplicación de las técnicas de clasificación

Excepto para el análisis de discriminante lineal que se utiliza el paquete *MASS*, se ha priorizado el uso del paquete *caret* y sus funciones para la elaboración del código. El paquete *caret* es actualmente el más utilizado en proyectos de *machine learning*, este paquete proporciona una visión consistente y unificada para todos los algoritmos. Las funciones utilizadas para la clasificación se enuncian a continuación.

**train:** esta función establece una cuadrícula para la sintonización de los parámetros para un amplio conjunto de modelos de clasificación y regresión. Se adapta a cada modelo y calcula una medida de rendimiento basada en el remuestreo.

**Train (y, data, method = M , tuneGrid = P)**

*y*: un vector de factores o numérico que contiene la clase correspondiente a cada observación.

*data*: matriz, data frame u otro tipo de dato donde las observaciones están ordenadas en filas y las características en columnas.

*method*: cadena de texto donde se indica el modelo de clasificación o regresión utilizado. Los algoritmos empleados son: knn, svmLin, svmLog, rf.

*tuneGrid*: un conjunto de datos con posibles valores para los parámetros. Las columnas poseen el mismo nombre que los parámetros

**predict:** es una función genérica para calcular predicciones de los resultados de varios modelos de clasificación o regresión. La función depende de la clase *class*.

**predict (object,...)**

*object*: es un objeto de tipo modelo sobre el que se realizará la predicción.

*test*: se pasa como argumento de la función la matriz de datos de test para la realización de la predicción.

**confusionMatrix**: calcula una matriz de confusión de las clases observadas y predichas con sus estadísticas asociadas.

**confusionMatrix (data, ...)**

*data*: vector de factores predichos

*test\$merger*: vector de factores reales.

**lda**: crea un modelo de análisis de discriminante linear.

**lda (y, data)**

*y*: un vector de factores o numérico que contiene la clase correspondiente a cada observación.

*data*: matriz, data frame u otro tipo de dato donde las observaciones están ordenadas en filas y las características en columnas.

- *Análisis de discriminante linear*:

***#creacion modelo***

**lda\_fit = lda(merger ~., data=train)**

***#predecimos para test***

```
test_pred = predict(lda_fit, , newdata = test)
#creamos la matriz de confusion
cm=confusionMatrix(test_pred,test$merger)
```

El funcionamiento de lda es bastante simple por lo tanto no requiere de ninguna adaptación particular.

- *K- Vecinos más cercanos:*

```
#creacion modelo
knn_fit= train(merger ~., data=train, method = "knn",
tuneGrid=expand.grid(k=seq(1,61,by=2)))
#predecimos para test
test_pred = predict(knn_fit, newdata = test)
#creamos la matriz de confusion
cm=confusionMatrix(test_pred, test$merger)
```

Para knn se contempla tuneGrid para determinar el valor del parámetro k, que indica el número de vecinos próximos que influirán en la clase de la observación a predecir. Para evitar empates en las votaciones el valor de la k siempre se considera un número impar que va desde 1 hasta 61 (raíz cuadrada del número de observaciones del conjunto de entrenamiento). El parámetro más repetido en la creación del modelo es k=61.

- *Máquinas de vector soporte:*

***#Parametros***

```
gridCost <- expand.grid(cost = 10^(-3:4))
```

***#entrenamiento svmLin con 'caret' package***

```
svmLin_fit <- train(merger ~., data=train, method = "svmLinear",  
                    tuneGrid=gridCost)
```

***#predecimos para test***

```
test_pred = predict(svmLin_fit, newdata = test)
```

***#creamos la matriz de confusion***

```
cm=confusionMatrix(test_pred, test$merger)
```

En svm linear se contempla tuneGrid para determinar el valor del parámetro C, este es un parámetro que propone la idea de una frontera de decisión ‘suave’ en la que se permita el error de clasificación de las observaciones más cercanas a ella. El parámetro más repetido en la creación del modelo es C= 1000, como es normal toma valores considerablemente altos debido a las dificultades de modelado que presenta el conjunto de datos.

***#Parametros***

```
gridCostGamma <- expand.grid(sigma = 10^(-5:2), cost = 10^(-3:4))
```

***#entrenamiento svmLog con 'caret' package***

```
svmLog_fit <- train(merger ~., data=train, method = "svmRadial",  
                    tuneGrid=gridCostGamma)
```

***#predecimos para test***

```
test_pred = predict(svmLog_fit, newdata = test)
```

***#creamos la matriz de confusion***

```
cm=confusionMatrix(test_pred, test$merger)
```

En svm gaussiano se se contempla tuneGrid para determinar el valor del parámetro C y gamma. Gamma es usado como la medida de similitud entre dos puntos. De forma intuitiva, decimos que un gamma pequeño define una función Gaussiana con una larga varianza, por lo cual dos puntos pueden ser considerados similares aunque la distancia entre ambos sea considerable. El parámetro más repetido en la creación del modelo es C= 1000 y gamma=1.

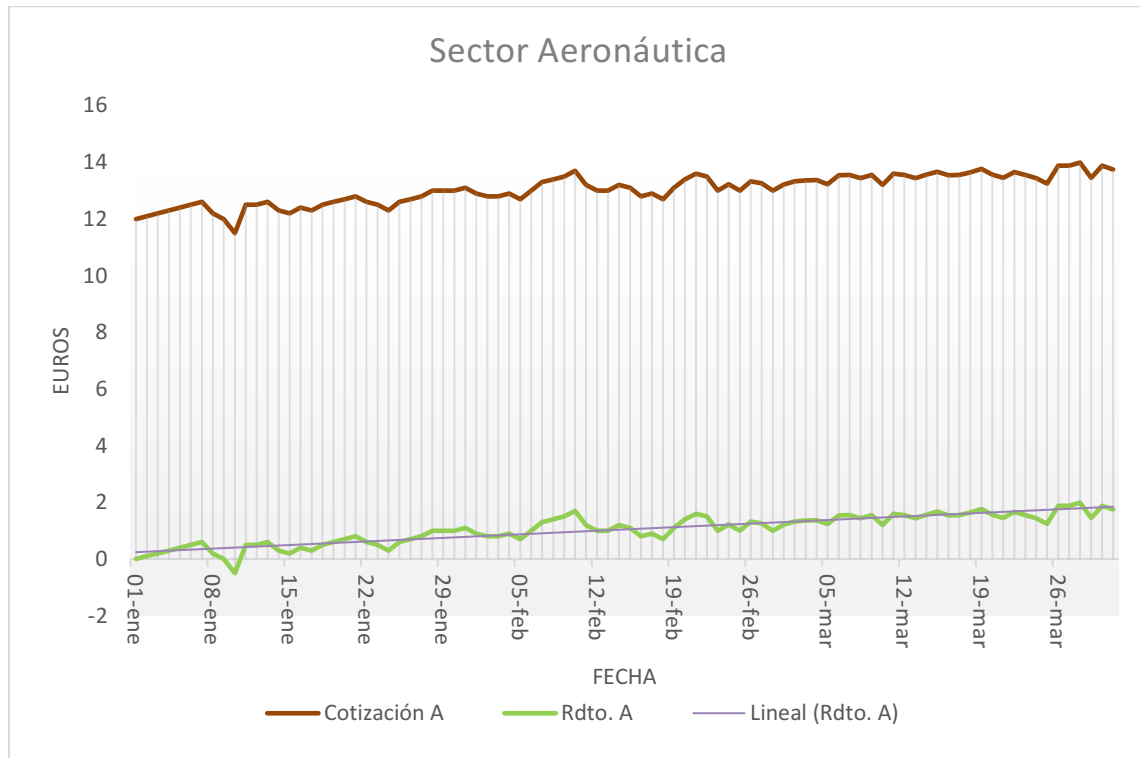
- *Random Forest:*

```
#entrenamiento rf con 'caret' package  
rf_fit = train(merger ~., data=train, method="rf", nodesize=1)  
#predecimos para test  
test_pred = predict(rf_fit, newdata = test)  
#creamos la matriz de confusion  
cm=confusionMatrix(test_pred, test$merger)
```

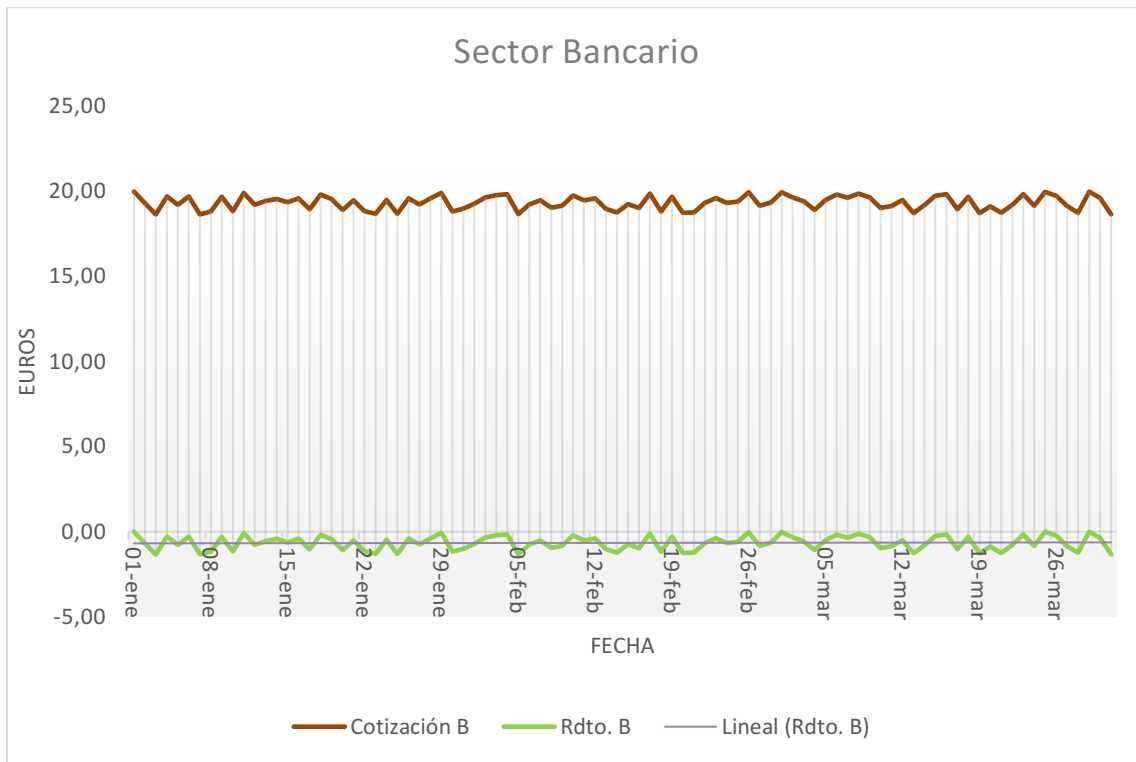
El algoritmo random forest hace por defecto una búsqueda del parámetro mtry que indica el número de variables consideradas en la creación de cada nodo del árbol. El parámetro mtry más repetido tiene el valor de 2.

## Apéndice 4: Cotizaciones simuladas

### *Estrategia en largo*

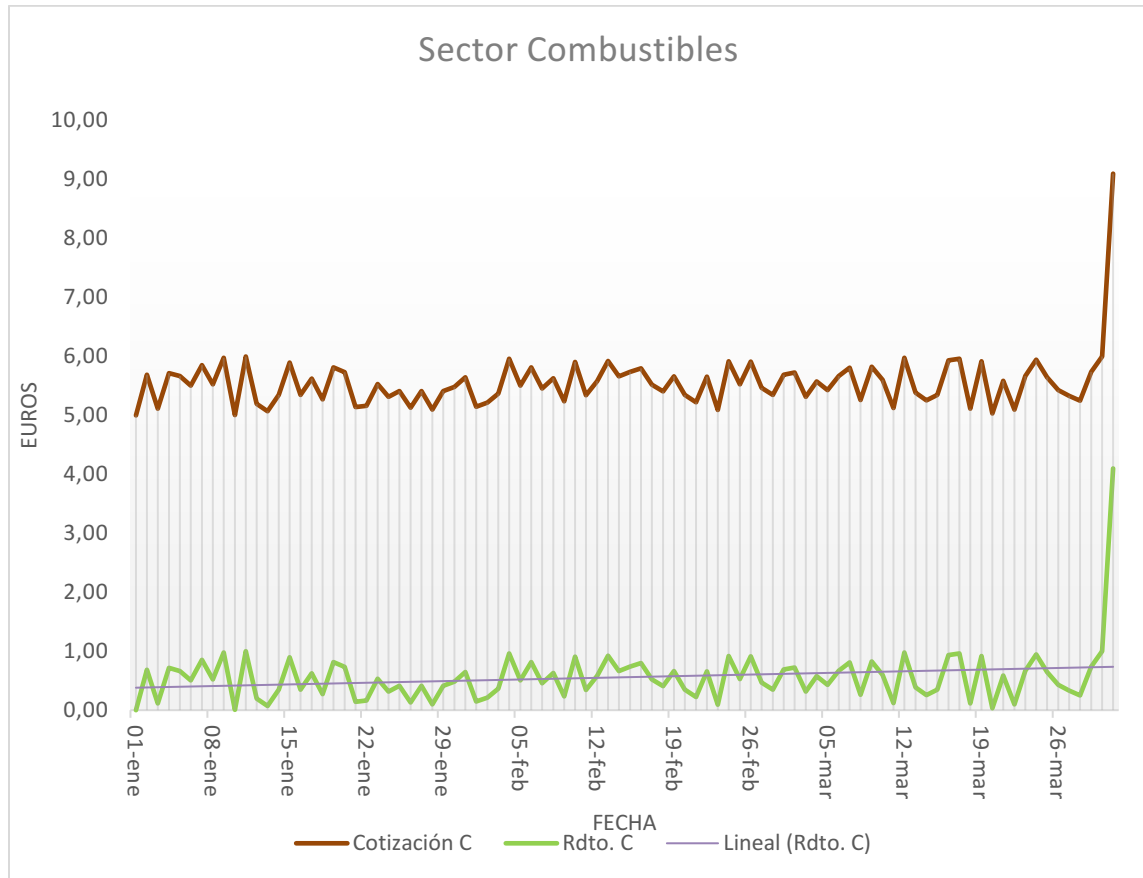


*Ilustración 12 – Arbitraje en largo empresa A. En esta gráfica se observa la evolución de las acciones para la empresa A durante el primer trimestre del año. Cotización de A indica el precio que posee A en el mercado. Rdto. A indica los beneficios potenciales que se obtendrían en el caso de vender la acción en ese instante. Lineal(Rdto. A) indica la tendencia de los beneficios potenciales para A*



*Ilustración 13 - Arbitraje en largo empresa B. En esta gráfica se observa la evolución de las acciones para la empresa B durante el primer trimestre del año. Cotización de B indica el precio que posee B en el mercado. Rdto. B indica los beneficios potenciales que se obtendrían en el caso de vender la acción en ese instante. Lineal(Rdto. B) indica la tendencia de los beneficios potenciales para B*

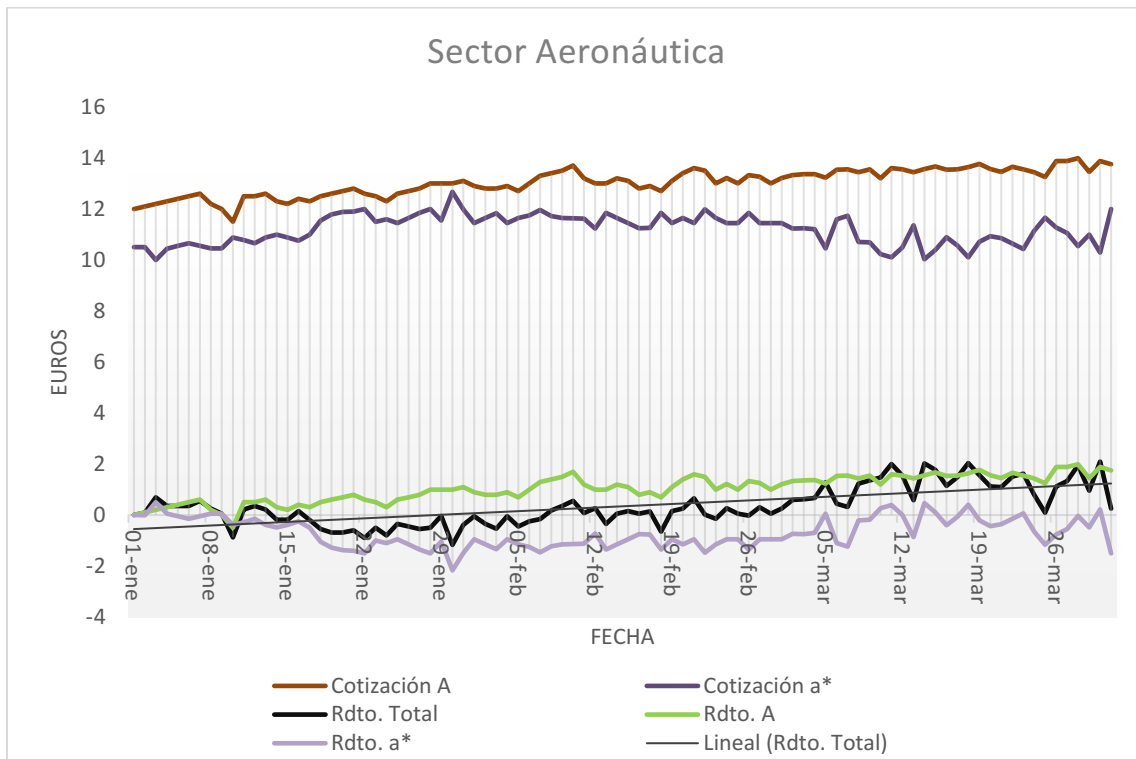
## APENDICE



*Ilustración 14 - Arbitraje en largo empresa C, empresa absorbida. En esta gráfica se observa la evolución de las acciones para la empresa C durante el primer trimestre del año. Cotización de C indica el precio que posee C en el mercado. Rdto. C indica los beneficios potenciales que se obtendrían en el caso de vender la acción en ese instante. Lineal(Rdto. C) indica la tendencia de los beneficios potenciales para C*

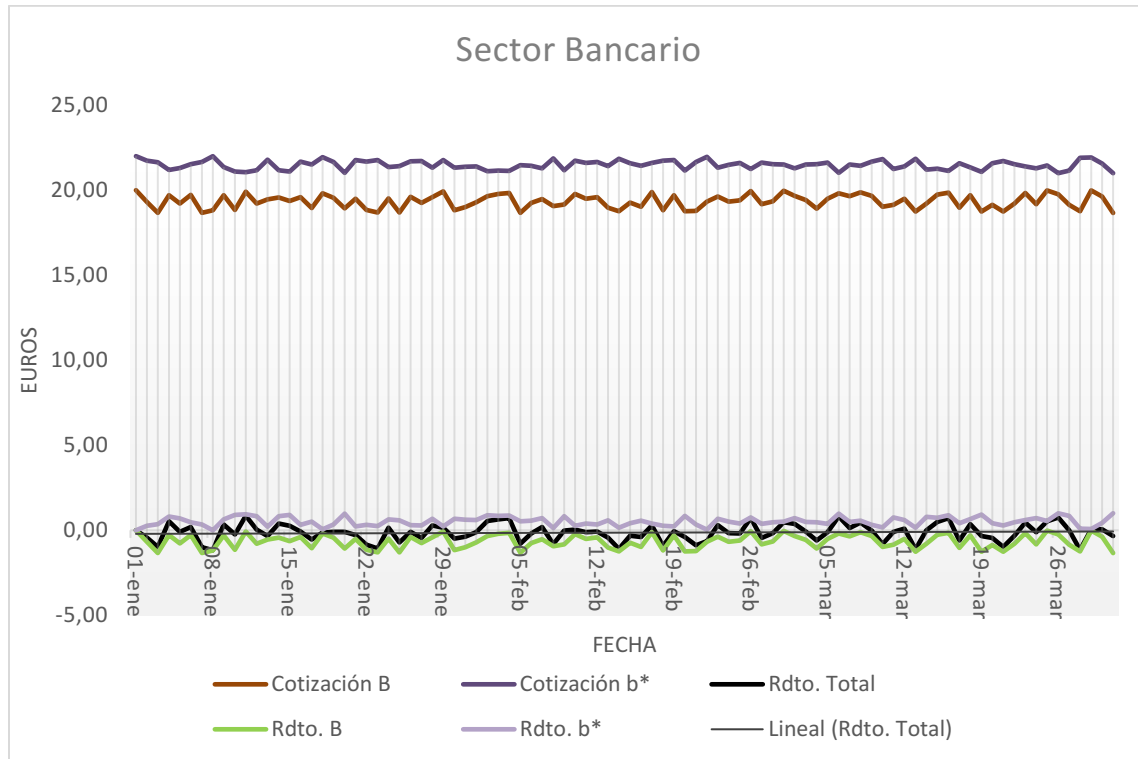


*Estrategia con empresas correlacionadas*

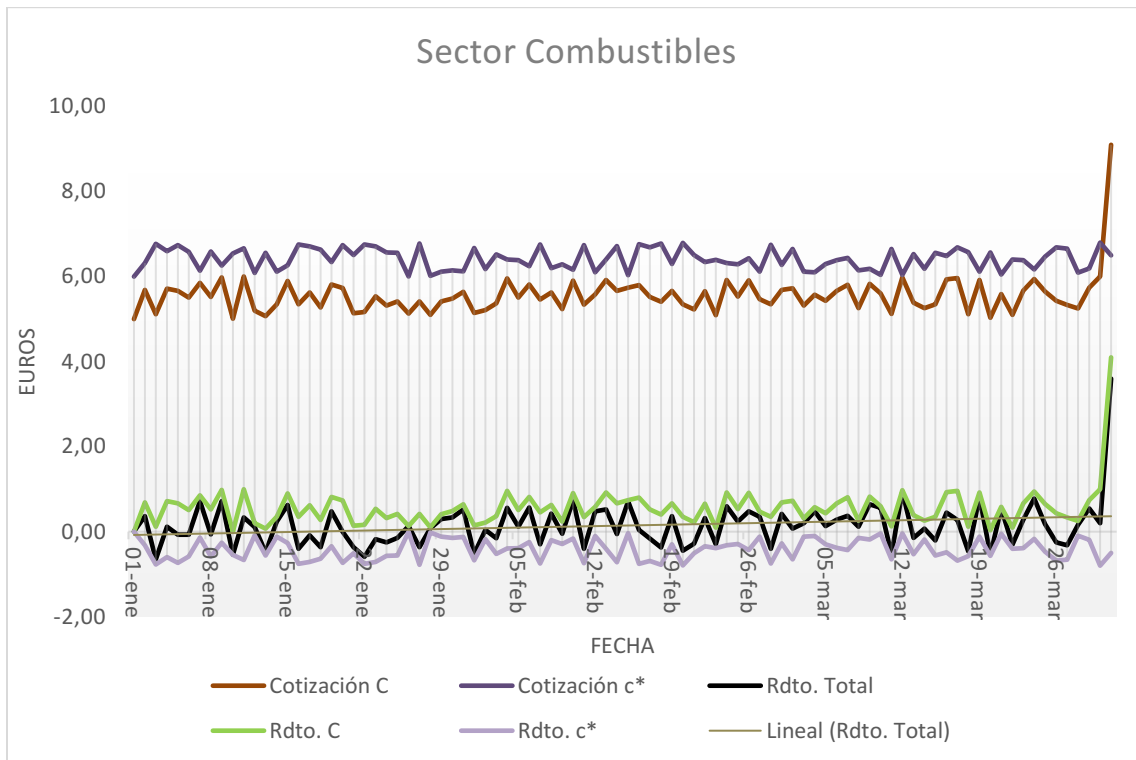


*Ilustración 15 – Estrategia con empresas correlacionadas empresa A. En esta gráfica se observa la evolución del precio de las acciones para la empresa A y una empresa correlacionada a ella a\* durante el primer trimestre del año. Cotización de A indica los beneficios que se obtendrán en el caso de vender la acción en ese instante. Cotización de a\* también nos informa sobre los beneficios al instante, pero al ser una estrategia en corto se comporta de forma opuesta a la tendencia de a\*. Rdto. Total se calcula a través de la suma de los beneficios generados por el Rdto. De A y de a\* y Lineal (Rdto. Total) es la tendencia de estos beneficios potenciales.*

## APENDICE



*Ilustración 16 - Estrategia con empresas correlacionadas empresa B. En esta gráfica se observa la evolución del precio de las acciones para la empresa B y una empresa correlacionada a ella b\* durante el primer trimestre del año. Cotización de B indica los beneficios que se obtendrán en el caso de vender la acción en ese instante. Cotización de b\* también nos informa sobre los beneficios al instante, pero al ser una estrategia en corto se comporta de forma opuesta a la tendencia de b\*. Rdto. Total se calcula a través de la suma de los beneficios generados por el Rdto. De B y de b\* y Lineal (Rdto. Total) es la tendencia de estos beneficios potenciales.*



*Ilustración 17 - Estrategia con empresas correlacionadas empresa C, empresa absorbida. En esta gráfica se observa la evolución del precio de las acciones para la empresa C y una empresa correlacionada a ella c\* durante el primer trimestre del año. Cotización de C indica los beneficios que se obtendrán en el caso de vender la acción en ese instante. Cotización de c\* también nos informa sobre los beneficios al instante, pero al ser una estrategia en corto se comporta de forma opuesta a la tendencia de c\*. Rdto. Total se calcula a través de la suma de los beneficios generados por el Rdto. De C y de c\* y Lineal (Rdto. Total) es la tendencia de estos beneficios potenciales.*

# Anexo

## Anexo 1: Palepu (1986)

Estimates of logit acquisition likelihood models. <sup>a</sup>					
Variables <sup>b</sup>	Expected sign	Estimates <sup>c,d</sup>			
		Model 1	Model 2	Model 3	Model 4
Average excess return	—	—1.332 (—2.53) <sup>e</sup>	—1.338 (—2.50) <sup>e</sup>		
Return on equity	—			0.003 (0.086)	0.005 (0.11)
Growth-resource dummy	+	0.5467 (2.47) <sup>e</sup>	0.4432 <sup>f</sup> (1.86) <sup>h</sup>	0.4616 (2.32) <sup>e</sup>	0.4024 (1.88) <sup>h</sup>
Growth			—0.0245 (—2.65) <sup>e</sup>		—0.0261 (—3.18) <sup>e</sup>
Liquidity			—0.005 (—0.49)		—0.008 (—0.85)
Leverage			—0.0035 (—2.07) <sup>e</sup>		—0.0034 (—2.17) <sup>e</sup>
Industry dummy	+	—0.7067 (—2.97) <sup>e</sup>	—0.6900 (—2.86) <sup>e</sup>	—0.5802 (—2.75) <sup>e</sup>	—0.5608 (—2.61) <sup>e</sup>
Size	—	—0.0005 (—2.61) <sup>e</sup>	—0.0005 (—2.62) <sup>e</sup>	—0.0004 (—2.52) <sup>e</sup>	—0.0004 (—2.63) <sup>e</sup>
Market-to-book ratio	—	—0.0044 (—0.17)	0.0117 (0.33)	—0.0051 (—0.2)	0.0126 (0.36)
Price-earnings ratio	—	0.0065 (0.78)	0.0099 (1.08)	0.0031 (0.51)	0.0041 (0.636)
Constant		—2.1048 (—2.49) <sup>e</sup>	—2.1096 (—2.45) <sup>e</sup>	—2.1533 (—3.35) <sup>e</sup>	—2.1898 (—3.47) <sup>e</sup>
Likelihood ratio index <sup>c</sup>		0.1010	0.1245	0.0695	0.0979
Likelihood ratio statistic <sup>f</sup>		58.65	72.32	47.78	67.29

<sup>a</sup>From a sample of 163 target firms that were acquired during the period 1971–1979 and 256 non-targets that were not taken over as of 1979. All the firms (1) belong to mining manufacturing industries, (2) were listed on the New York or the American stock exchange, and (3) have data on COMPUSTAT and CRSP tapes. For more details on the sample, see section 3.3.

<sup>b</sup>The independent variables are measured as of the end of the fiscal year prior to the year of takeover for targets and as of the end of the fiscal year prior to 1979 for non-targets. For a complete description of how these are computed, see the appendix.

<sup>c</sup>Four different versions of the model are estimated. Model 1 consists of six variables corresponding to the six hypotheses in table 1. Model 2 is a re-estimation of model 1 with the three additional variables growth, liquidity and leverage. Models 3 and 4 are re-estimations of models 1 and 2, respectively, with accounting return on equity replacing average excess return, a market performance measure. The constant term in all the four models is corrected for the sampling bias.

<sup>d</sup>The *t*-statistic, computed to test the null hypothesis that the estimated coefficient is equal to zero, is shown in parentheses for each coefficient estimate.

<sup>e</sup>The log likelihood ratio index is defined as  $(1 - \log \text{likelihood at convergence}) / \log \text{likelihood at zero}$ . It is similar to the  $R^2$  statistic in the case of a multiple regression model and provides an indication of the logit model's explanatory power.

<sup>f</sup>The likelihood ratio statistic is computed to test the hypothesis that all the parameters in the model are simultaneously equal to zero. Under this null hypothesis, the statistic has an asymptotic distribution which is a chi-square with the degrees of freedom equalling the number of parameters in the model. The statistic is significant at the 0.01 level for all the models.

<sup>g</sup>Significant at the 0.05 level, two-tailed test.

<sup>h</sup>Significant at the 0.10 level, two-tailed test.

Tabla 25 – Resultados para (Palepu, 1986)

### **Abstract**

This TFG studies the prediction of company mergers and acquisitions through machine learning techniques. The prediction of this kind of operations with these machine learning techniques becomes an interesting problem since it has not been yet approached properly and therefore is subject to improvement. To do the forecast we dispose of a set of observations with financial data of American companies between 1994 and 2014. The study is carried out with different classifiers in order to evaluate the results from different approaches. The analysis shows quite pessimistic results in all the classification algorithms; however, not only the relevant investment strategies are studied but also how to obtain positive returns from a merger or acquisition.

**Palabras clave:** Acquisition, Machine Learning, Arbitrage, Classification, Merge, R, Observation, Panel, Forecast.

ENGLISH VERSION

## **Introduction**

### **Chapter**

#### **Motivation, goals and limitations**

Corporate takeovers are represent, even more, an important factor for business developed in a competitive environment and it is a strong tool that companies utilize to maintain or even expand their market power and the competitive advantages.

The advances of computation and information transmission techniques has generated the creation of huge information. Even so, much of this information has not been processed yet and continues stored, without being used, remaining as raw data. As there is a lot of stored data, the potential information the can be obtained from this data is potentially high. Through the use of machine learning techniques data can be modeled and analysed to detect patterns and therefore gather useful information.

Shown by different studies Jarrell & Poulsen (1989) and Andrade, Mitchell, & Stafford (2001) it is well known that shareholders of the acquired companies receive high profits after the bid announcement for the company. Therefore, finding ways to identify these companies before an bid offer is announced is a problem of great interest to the financial industry. However, this is a complex problem that still has no solution.

Several studies have proposed prediction models public information based for takeover operations, mainly oriented to companies in the US market Palepu (1986), Meador, Pamela, & Gayle (1996), Barnes (2000), Ragothaman, S., Naik , B., & Ramakrishnan

(2003). Most of these models use as predictive techniques the factor analysis, linear discriminant analysis or logarithmic regression models.

The aim of this TFG is to identify influence factors and verify their viability for merger operations in the US market. To reach this, machine learning techniques that will predict from financial data if a takeover operation has been accomplished at the end of a financial quarter will be implemented. Based on results, the potential to generate positive turnovers applying suitable investment strategies for M&A will be analysed.

This TFG attempts to contribute to the existing studies in three points. First, the use of machine learning techniques for the prediction of mergers and acquisitions. Second, the feasibility of using fully financial and public data. Third, the applicability of the obtained results to a real scenario.

Nevertheless, there are different limitations for the realization of a single prediction model, valid for all companies. Machine learning techniques will only process financial data that, while providing significant value to prediction, does not cover all the reason why takeovers occur. There are other variables, not considered in the study such as market variables (Meador et al., 1996), macroeconomic variables (Rossi & Volpin, 2004), technological variables (Yang, 2014) and variables related to the control structure (Ambrose & Megginson, 2011) which have also influences on the decision of a merger operation. Related to the data used in the study, although they have been gathered from reliable sources (*Compustat* and *SDC Plantinum*), it has not been possible to verify the validity of all of them, so they have been assumed to be correct in order to develop the prediction.



## **Document structure**

Once exposed the main goals that the development of this work intends, the structure of the document can be divided into three blocks. These blocks are preceded by the introductions and followed by the section of conclusions and further research.

The first block (Chapter 2) contains the theoretical foundations of mergers and acquisitions. It presents the reasons that lead two companies to a merge, the influences that macroeconomic tendencies generate in this type of operations and a summary of the prediction models applied previously. It also presents in a general way what is machine learning and its application areas.

The second block (Chapter 3,4,5,6) describes the elaboration of the prediction empirical study. Applied methodology, used data, classification techniques and evaluation of the model are sections of this block. It also describes sections related to the theoretical foundations in order to understand the decisions taken.

ENGLISH VERSION

## **Literature review**

### **Chapter**

#### **Definitions and causes for M&A**

The increase rate of M&A operations in the United States in recent years has been remarkable. Brealey, Myers & Allen (2013) states:

a record year for mergers, companies were involved in deals totaling nearly \$1.5 trillion in the USA and a further \$0.5 trillion in Europe. Asian companies, always laggards in the urge to merge, accounted for just \$59 billion of deals. (p.824)

The number of such operations has increased and by 2015 has been \$5.0 trillion. During this periods of intense activity, management spends significant amounts of time trying to foresee if one company will acquire another.

We commonly refer to the terms of merger or acquisition in almost the same way, but these are different. A merger arises when two companies decide to join to form a single company. On the other hand, an acquisition, generally exists when a large company decides to buy a smaller company. As they are different operations, they follow different stages and restrictions, but the final result is practically the same on what the behaviour of the company in the stock market concerns. Therefore, throughout this study no difference will be made between these two terms, we will group them under the term of merger operations and we analyse them under the assumptions set forth by Toberfte & Curtis (2013) that uses the term merger for any corporate acquisition or control through the purchase of shares of

an objective company, regardless of whether the transaction was made through an agreement or a public offer.

Merger operations mainly focus all their appeal on the economic result of combining the two companies, but they are also related to which of the two manages the resulting company. In any merger transaction it is considered that there is a leading firm called acquirer and another considered as target, where usually the management of the acquired firm leaves their control positions after the operation.

The M&A market is currently considered as part of a larger market for corporate control that includes privatizations, nationalizations and purchases of companies in competition from creditors, among others. These operations, which arise for different reasons and have different consequences, will not be analyzed during the study.

### **Machine learning**

Machine Learning is a subfield of computer science and a branch of artificial intelligence that aims to develop techniques that allow computers to learn and be without being explicitly programmed (Rouse,2017). Based on provided information as examples, the machine learning techniques are able to create programs capable of generalizing behaviours.

Machine Learning techniques try to solve problems of the statistics domain that present a high computational complexity, in this way arise different disciplines of this matter like the data analysis or the study of computational complexity.

In order to understand where machine learning comes within the area of Artificial Intelligence, it is necessary firstly to understand the concepts of agent, environment, sensor and actuator. Russel & Norvig (2010): “An agent is anything that can be seen or perceived in your environment or through sensor and that interacts with that environment through actuators”. An example applied to robotics would be the case of a robot that has cameras and infrared sensors as sensor and several motors as actuators and act in a certain location. Once these four concepts are established, the objective of artificial intelligence is to get the agent to behave rationally, which means that for each possible sequence of actions, the agent selects an action that maximizes its performance.

The agent's behaviour is affected by what it perceives in the environment through its sensor, this is called environment problem, This concept considers the problems that are presented to rational agents and that are independent to it, but to which it may provide the solutions [*Ilustración 2*].

After discussing the concepts of “environment problem” and “rational agent” we observe that different classifications can be made in artificial intelligence depending on these two concepts. In relation to the environment we could make classifications depending on their properties (totally or partially, observable, static or dynamic, deterministic or stochastic..) and in relation to the agents in how they have been programmed (utility, objectives, ...) But the classification that really interests us is the one given by the type of data obtained from the environment.

In this study machine learning techniques that are implemented with supervised learning algorithms with a classification-oriented approach will be used. Supervised learning consists of making a correct prediction of an input data set, with an algorithm that has previously been trained by a set of training data with a valid output. Predictions can be a

ENGLISH VERSION

continuous value (as in regression problems) or a class label (classification) which has discrete values.

## Methodology

### Chapter

The methodology used in this project is based on CRISP-DM (Cross Industry Standard Process for Data Mining) (Marbán, Mariscal & Segovia, 2009) because it has a global approach to solve machine learning projects [*Ilustración 3*]. From this methodology we will create a framework where the different parts of the experiment will be established. Due to this is a methodology used in data mining and knowledge discovery processes and the problem to be solved is a *classification problem*, some parts will be adapted to the used of supervised learning techniques.

The stages of this methodology will be briefly described below. Similar methodologies and more details of each stage can be found in any monograph about the subject (Marbán et al., 2009).

#### Business understanding

This phase focuses on understanding the objectives and requirements of the project from a business perspective. In this case, the creation of a portfolio of companies stocks that potentially can be acquired at the end of the quarter.

Subsequently, it is carried out an evaluation of the problem to be solved and the study is presented from a point of view oriented to machine learning. The objectives to be reached are determined and a plan to obtain the desired result is finally created.

## **Data understanding**

The data understanding phase begins with the data collection from the information sources. It continues with a preliminary analysis, which consists on obtaining a first idea of the type of data that is available, its description and common features. Subsequently a dependent analysis variable is performed to observe its characteristics and to find out which are the factors that make influence on it. A quality analysis of the data is also performed, observing possible atypical data, observations without information and problems related to quality for future correction. Finally, a simple descriptive analysis of the most influential variables in the model is carried out with the purpose of discovering underlying relationships in the data set.

## **Data preparation**

Data preparation focuses on the adaptation of the initial data so that, it can later be used in an appropriate way. This adaptation is done based on the data understanding section leading as a result the final data set. The process of data preparation does not have to follow an established order since it is carried out simultaneously to the elaboration of the model. This phase includes the selection, deployment, construction, integration and formatting of the data.

## **Modeling**

In this phase modelling techniques are selected according to the type of problem to be solved. In this case, as it is a classification problem, suitable techniques to solve this problem will be used. The development of the classification and evaluation model are the core part of this phase.



Due to some techniques have specific requirements in the structure of the data, this phase does not have a pre-set order, and in some moments it would be necessary to go back to previous steps to adapt the data to the techniques used at each moment.

### **Evaluation**

Once results of the empirical experiment are obtained, they are evaluated from a data analysis perspective. After this analysis, a whole process review will be made to later define the following steps, it is important to carry out a thorough evaluation of the results, model and the taken decision to ensure that result are able to meet the objectives of the business problem. At the end of this phase, an overall view of the strengths and weaknesses of the applied techniques and possible improvements are made.

### **Deployment**

In this research, deployment phase indicates the way in which the prediction model would be applied and how it would be included in the desired business objectives.

ENGLISH VERSION

## **Conclusions and future work**

### **Chapter**

#### **Conclusions**

The classification problem proposed is an interesting problem considering that from the point of view of machine learning takeover operations have not been addressed yet. The conclusions obtained from this study are of two types: related to machine learning and financial topics.

Due to the computational complexity it entails, it is a very difficult problem to solve and it poses many difficulties because the data that best predict it are not clearly known.

The problem presented an extreme imbalance in the data, which, for machine learning techniques represents a problem of lack of performance.

With the use of financial variables for prediction (atq, ceqq, dltqq, niq, saleq, per, roe, leverage) the standard methods have not get the expected performance, although some positive signs can be observed. The error in the classification is smaller than the one of a random classification, therefore, either improving the data or improving the methodology could improve the results.

The information obtained from the prediction would not be efficient in an extrapolation to the real world applied to the management of a stock portfolio.

### **Future work**

Below, some lines of future work with which to improve the model presented in the study will be described.

First, from the current data, other indicators could be created. Through an analysis of attributes we could eliminate the variables that do not contribute information to the model and create others based on financial ratios that the literature endorses as significant. A configuration of the panel with variables where the evolution of the company in the time study could contribute much information for the classification.

Second, with reference to the type of training, other classification algorithms such as neural networks could be considered and a more exhaustive optimization of those already used.

Third, adding to the panel of market data, technological or macroeconomic would provide a more global view of the problem, since it is not only the economic reasons that generate this type of operations.

Fourth, the search for clusters or groups of companies with similar characteristics to be able to perform the classification within these subgroups.

Finally, apart from just studying the potential target companies, also emphasize the potential companies that can make an offer. In this way if we can see similarities between companies of which one is potential target and another potential buyer we would have more information about a possible absorption.